THE POLYKETIDE ORIGINS OF CANNABINOIDS IN *CANNABIS SATIVA*

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Biology

University of Saskatchewan

Saskatoon

By

STEVE JOSEPH GAGNE

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Biology

University of Saskatchewan

Saskatoon, Saskatchewan, S7N 5E2

ABSTRACT

Phytocannabinoids are the active substances responsible for the medicinal and psychotropic effects of *Cannabis sativa*. Although the bioactivity of cannabis and its preparations have been known for millennia, several steps in the biosynthetic pathway leading to phytocannabinoids remain unclear. Phytocannabinoids are prenylated resorcylic acids which are formed in specialized plant organs called glandular trichomes. Following the analysis of a pre-generated cannabis trichome cDNA library, a type III polyketide synthase (tetraketide synthase; TKS) was identified and assayed, yielding three major compounds, hexanoyl triacetic acid lactone (HTAL), pentyl diacetic acid lactone (PDAL), and olivetol, yet no resorcylic acid was detected. This lack of resorcylic acid in enzyme assays has instigated the characterization of TKS and a search for putative cyclases in the cannabis trichome cDNA library, and involved protein pulldown, co-immunoprecipitation, and co-assay experiments. These experiments led to the discovery of a novel polyketide cyclase protein named olivetolic acid cyclase (OAC) responsible for the proper cyclization of a polyketide intermediate produced by TKS. This thesis shows that TKS assays conducted with OAC produce olivetolic acid (OA), an intermediate required during the biosynthesis of cannabinoids. The TKS/OAC spatial relationship was also investigated following the creation of fluorescent fusion proteins which show that the enzymes co-localized *in vivo* when viewed with confocal microscopy. Furthermore, yeast two-hybrid assays using TKS and OAC were performed to establish whether the enzymes physically interact. Finally, an attempt to determine the responsible amino acids involved in OAC's mechanism was conducted by comparing the activity of single point OAC mutants with the wild-type OAC. Based on the available data, mechanisms for the production of HTAL, PDAL, olivetol, and OA are proposed.

ACKNOWLEDGEMENTS

Although the research presented in this dissertation is my work, this project would not have been possible without the involvement of many people who have ceaselessly provided their valuable support. I would primarily like to thank my supervisor, Dr. Jonathan Page, who offered me the opportunity to tackle the cannabinoid conundrum, and continued to provide encouragement and guidance throughout each process of this work. I'd like to also thank Jon for the important bits of advice he has passed on to me with regard to professionalism and work flow, which will continue to assist me in my future career, whether it is in industry, academia, or government settings. I also warmly thank each member of my committee who have provided important comments that helped to shape this thesis to its current state; each have generously provided flexibility with their time and personally contributed important input during the course of this project; thank you Dr. Gordon Gray, Dr. Chris Todd, and Dr. David Palmer. I also thank the external examiner, Dr. Reinhart Jetter, who has also contributed a lot of significant input during the final stages of my dissertation preparation.

I would like to express my sincere gratitude to each lab member, whether past or present, which I had the pleasure to work with during these last few years. This includes Dr. Jake Stout, Dr. Shawn Clark, Enwu Liu, David Konkin, Sandra Polvi, Dr. Zakia Boubakir, Shawn Whitfield, and Nikki Theaker. Aside from providing their invaluable technical expertise, exemplary teaching and guidance, and stellar participation and inspiration, I am deeply appreciative for their friendship, openness, and for being the best lab mates one could ask for.

I would also like to thank each and every individual that I had the opportunity to meet at the University of Saskatchewan and the Plant Biotechnology Institute. The environment that

these people have collectively created is one of highest integrity and good cheer. A special thanks to Dr. Patrick Covello, Darwin Reed, Stephen Ambrose, Dr. Shawn Gibson, Sharla Lozinsky, Dr. Julien Cotelesage, Deidre Wasyliw, Brenda Haug, Dr. Leonid Akhov, Ioannis Mavraganis, and Aaron Karapinka; your precious words have coaxed me to the finish line.

I acknowledge the financial support provided to me by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Department of Biology, and the University of Saskatchewan.

Lastly, but most importantly, I would like to thank my best friend and wife, Christine, as well as my children, Spring and Sethia Om, for I know well enough that without their enduring love, kindness, and patience, I would not have completed one word of this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3-AT	3-amino-1,2,4-triazole
6-DDHK	6-deoxydihydrokalafungin
AA	amino acid
Ab	Antibody
ABM	antibiotic biosynthetic monooxygenase
ACP	acyl carrier protein
AD	GAL4 Activating Domain
ADA	N-(2-Acetamido)iminodiacetic acid
Amp	Ampicillin
Amp^R	Ampicillin resistance marker
*At*HS1	*Arabidopsis thaliana* heat stable protein
BDAL	butyryl diacetic acid lactone
BP	before present
BRI	Biotechnology Research Institute
BTAL	butyryl triacetic acid lactone
B-TKS	Talon-bead affixed recombinant TKS
C4H	cinnamate 4-hydroxylase
CBCA	cannabichromenic acid
CBD	cannabidiol
CBDA	cannabidiolic acid
CBG	cannabigerol

CBGA	cannabigerolic acid
CBN	cannabinol
CDD	Conserved Domain Database
CHI	chalcone isomerase
CHR	chalcone reductase
CHS	chalcone synthase
CNBR	cyanogen bromide activated Sepharose 4B
CoA	Coenzyme A
D4R	dihydroflavonol 4-reductase
DABB	dimeric α + β barrel
DBD	GAL4 DNA binding domain
DHK	dihydrokalafungin
DMAPP	dimethylallyldiphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
Gen	gentamycin
Gen^R	gentamycin resistance marker
GOT	geranylpyrophosphate:olivetolate transferase
GPP	geranyl diphosphate
HDAL	heptyl diacetic acid lactone
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HPLC	high performance liquid chromatography

HTAL	hexanoyl triacetic acid lactone
IMAC	immobilized metal chromatography
IPP	isopentenyl diphosphate
JTT	Jones-Taylor-Thornton
Kan	Kanamycin
Kan^R	Kanamycin resistance marker
LCMS	liquid chromatography mass spectrometry
MCS	malonyl-CoA synthethase
MLMI	methylmuconolactone methyl isomerase
MLP	Betv1-like (major latex) protein
MOPSO	3-(N-morpholinyl)-2-hydroxypropanesulfonic acid
MRM	multiple reaction monitoring
MW	molecular weight
MWCO	molecular weight cut-off
NNI	Nearest-Neighbor-Interchange
NRC	National Research Council Canada
OA	olivetolic acid
OAC	olivetolic acid cyclase
OAC:YFP	OAC-yellow fluorescent protein
OLS	olivetol synthase
OMT-1	*O*-methyltransferase
OTAL	octanoyl triacetic acid lactone

PAL	phenylalanine ammonia-lyase
PBI	Plant Biotechnology Institute
PDA	photo array detector
PDAL	pentyl diacetic acid lactone
PDB	protein databank
PKC	polyketide cyclase
PKS	polyketide synthase
*Pt*SP1	*Populus tremula* stable protein
PVPP	polyvinylpolypyrrolidone
Rmsd	root mean square deviation
RT	room temperature
RT-PCR	reverse transcriptase-polymerase chain reaction
SCOP	Structural Classification of Proteins database
SDM	site-directed mutagenesis
Spc	spectinomycin
Spc^R	spectinomycin resistance marker
SRABB	stress responsive alpha beta barrel protein
STCS	stilbene carboxylate synthase
S-TKS	soluble recombinant TKS
STS	stilbene synthase
TAL	triacetic acid lactone
THC	Δ^9-tetrahydrocannabinol

THCA	tetrahydrocannabinolic acid
TKS	tetraketide synthase
TKS:CFP	TKS-cyan fluorescent protein
TLC	thin layer chromatography
TRIS	2-amino-2-hydroxymethyl-propane-1,3-diol
TTAL	tetraacetic acid lactone
Uniprot	Universal Protein Resource database
VPS	valerophenone synthase
wtOAC	wild-type OAC
X-gal	5-bromo-4-chloro-3-indolyl- beta-D-galactopyranoside
Y2H	yeast two-hybrid
βME	β-mercaptoethanol

1. INTRODUCTION

Cannabis sativa is one of the most recognizable plants in our society having been harvested since antiquity as food and fibre, and due to its unique pharmaceutical properties. Due to its medicinal and psychotropic effects, the plant has attracted substantial attention, and is the source of controversy and close scrutiny by the international community. Yet, despite its importance, the legal status of cannabis has impeded its research, and the biosynthesis of the compounds known as cannabinoids, which are exclusively produced in cannabis, has remained unclear. It is the purpose of this dissertation to provide insight into a key step in the cannabinoid biosynthetic pathway, and to provide evidence that a type III polyketide synthase (PKS) and a dimeric $\alpha + \beta$ barrel (DABB) polyketide cyclase are both responsible and required for olivetolic acid (OA) biosynthesis. Tetraketide synthase (TKS) is a type III PKS initially identified from a cannabis trichome-specific cDNA library. Although it was anticipated that TKS was responsible for the biosynthesis of OA, a precursor involved in cannabinoid biosynthesis, the enzyme cannot independently produce OA, and instead yields a decarboxylated analog. This lack of OA-forming activity prompted an investigation into TKS, which eventually led to the discovery of a DABB protein called olivetolic acid cyclase (OAC).

Formally, this dissertation describes the initial characterization of TKS and the steps that resulted in the identification of OAC (Chapter 3), followed by a report on the *in vitro* optimization of the OA biosynthetic reaction (Chapter 4). Subcellular localization experiments and yeast two-hybrid assays were employed to determine if TKS and OAC physically interact, and progressed to the use of protein pulldowns and micro-dialysis experiments to further probe

the nature of the TKS/OAC reaction (Chapter 5). Furthermore, an attempt to determine which amino acids are involved in the OAC mechanism was conducted by monitoring and comparing the activity of 15 single point OAC mutants alongside the wild-type OAC (Chapter 6). Observations made within the course of this work support the hypothesis that OA biosynthesis requires 1) TKS to generate a labile tetraketide intermediate and 2) OAC for the intermediate's proper cyclization. This led to the MS/MS screening of TKS/OAC reactions for tetraketo-CoA linked analytes, including a theoretical short-lived polyketide intermediate (Chapter 6).

1.1 Research Goals and Objectives

The initial goal of this research was to determine why OA was not being produced from TKS. This required the characterization of TKS, the testing of alternative assay conditions to induce TKS to produce OA, and to identify putative cyclases identified in a cannabis trichome cDNA library (Stout et al. 2012). Once the polyketide cyclase OAC was identified, it was characterized and an attempt to delineate its underlying mechanism was conducted. This involved determining if TKS and OAC co-localized, physically interacted, and could produce OA when physically separated. Structure/function studies were also conducted in an attempt to discern amino acids involved in OAC's mechanism. It is also a goal of this research to implicate TKS and OAC as two principle enzymes required for the biosynthesis of OA, which are therefore components of the cannabinoid biosynthetic pathway. The information resulting from this thesis supports the hypothesis in favor of the existence of a transient linear tetraketide.

2. LITERATURE REVIEW

2.1 Overview of *Cannabis sativa* and Cannabinoids

Cannabis has been a multipurpose resource for thousands of years. Humans and other hominids have ingested its seeds since prehistoric times (Callaway, 2004). Archaeological evidence indicates that it was cultivated for its use as a textile in Neolithic China (*i.e.* 6500 before present; BP) (Fleming and Clarke, 1998). It is unknown when the psychoactive properties of cannabis were first discovered, although physical evidence suggests they were known as early as 2700 BP (Jiang et al., 2006; Russo et al., 2008). The earliest record of medicinal cannabis appeared in the first century in the pen-ts'ao ching, where it was recognized as a treatment for numerous ailments including rheumatism, constipation and malaria (Touwn, 1981). Cannabis was also used in western medicine during the early 20^{th} century, before it was outlawed and subsequently removed from the British and American pharmacopoeias in 1932 and 1944, respectively. The history of cannabis and its uses has been well reviewed by several authors (Abel, 1980; Fleming and Clarke, 1998; Russo et al., 2008; Zuardi, 2006).

The psychoactive properties of cannabis have long been the cause for scientific examination of the responsible biological compounds. The bioactive fraction from cannabis, which is now known to consist of many cannabinoids, was initially distilled from hemp resin and described as a red oil in 1896; the acetate crystals obtained from this oil were named cannabinol (CBN) and at the time were believed to be the psychoactive agent in cannabis (Wood et al., 1896). It was not until 35 years later that similar experiments were successfully reproduced, and that a partial structural elucidation of CBN was published (Cahn, 1932). In 1940, a second

cannabinoid named cannabidiol (CBD) was also isolated from red oil, but both CBN and CBD were found to be pharmacologically inert and lacked 'marihuana activity' (Adams et al., 1940a; Loewe, 1945). The psychoactive cannabinoid Δ^9-tetrahydrocannabinol (THC) was later discovered in 1940 as a product of *in vitro* CBD isomerisation, and was not isolated from cannabis, albeit in an impure form, until the following year (Adams et al., 1940b; Wollner et al., 1942). The isolation of pure THC from cannabis was eventually achieved in 1964, and its structure was determined in 1971 (Gaoni and Mechoulam, 1964; Gaoni and Mechoulam, 1971) (Figure 2.1).

Since the pioneering days of cannabis research, more than 100 cannabinoids exclusively produced in cannabis have been identified (Mechoulam and Hanus, 2000; Russo, 2011). Cannabinoids are a unique class of chemicals that are composed of terpene and phenolic moieties (*i.e.* terpenophenolics, or prenylated polyketides). They do not contain any nitrogen, although cannabis alkaloids have also been reported (Mechoulam, 1988). The structural elucidation and synthesis of THC allowed for the identification of their endogenous targets, the CB1 and CB2 receptors (Devane et al., 1988; Munro et al., 1993). In turn, the discovery of the cannabinoid receptors led to the identification of their endogenous ligands, arachidonoylethanolamide (anandamide) and 2-arachidonoylglycerol, and have since been recognized as components of the endocannabinoid system (Devane et al., 1992; Mechoulam et al., 1995) (Figure 2.1). Cannabinoids that are produced *in planta* are sometimes referred to as phytocannabinoids to distinguish them from endogenously produced CB1 and CB2 ligands, which are collectively known as endocannabinoids (Russo, 2011). Phytocannabinoids are produced as acids *in planta*

but are decarboxylated to their neutral form upon heating and storage (Kimura and Okamoto, 1970; Turner et al., 1980; Yamauchi et al., 1967).

2.1.1 The Biology of Cannabis

Cannabis is a dioecious annual herb that produces either male staminate or female pistillate flowers, although monoecious varieties are also known to exist (Figure 2.2). Staminate flowers develop five stamens with near sessile anthers that are surrounded by a calyx of five sepals. The male inflorescences are indeterminate racemes that form into a panicle arrangement of flowers that tend to bloom slightly earlier than their female counterparts. The pollen is released during the bloom and is normally dependent on wind to fertilize nearby female plants. Soon after staminate plants shed their pollen, they enter senescence and die. Female plants bear pistillate flowers which initially appear late during the long-day vegetative cycle as two elongate stigmas at branching nodes. During the short-day flowering cycle, the nodal bracts blossom and accumulate more stigmas, each surrounded by short lobed calyces that protect the developing fruits, known as achenes, or more commonly as seeds (Stearn, 1970). Although the existence of sexual dimorphism in cannabis is evident, the determining mechanism is not completely understood nor is it completely agreed upon. Some models for sexual determination have implied x/autosomal dosage mechanisms, although a growing body of evidence, including RAPD and cDNA-AFLP analysis suggests that a heterogametic X/Y system exists (Ainsworth, 2000; Mandolino et al., 2002; Moliterni et al., 2004).

Prior to 1998, cannabis belonged to the Urticales order, but since then this order has been embedded into the Rosales order (Sytsma et al., 2002). Cannabis is now circumscribed as an urticalean rosid (Rosales) and belongs to the Cannabaceae family (Soltis et al., 2011). The

Cannabaceae family is now recognized as *sensu lato* to include the Celtis genus (hackberry), or as *sensu stricto* as to only include the *Humulus* genus (hops). Analysis of cpDNA restriction sites support the relationship between Cannabaceae and Celtidaceae families (Song et al., 2001).

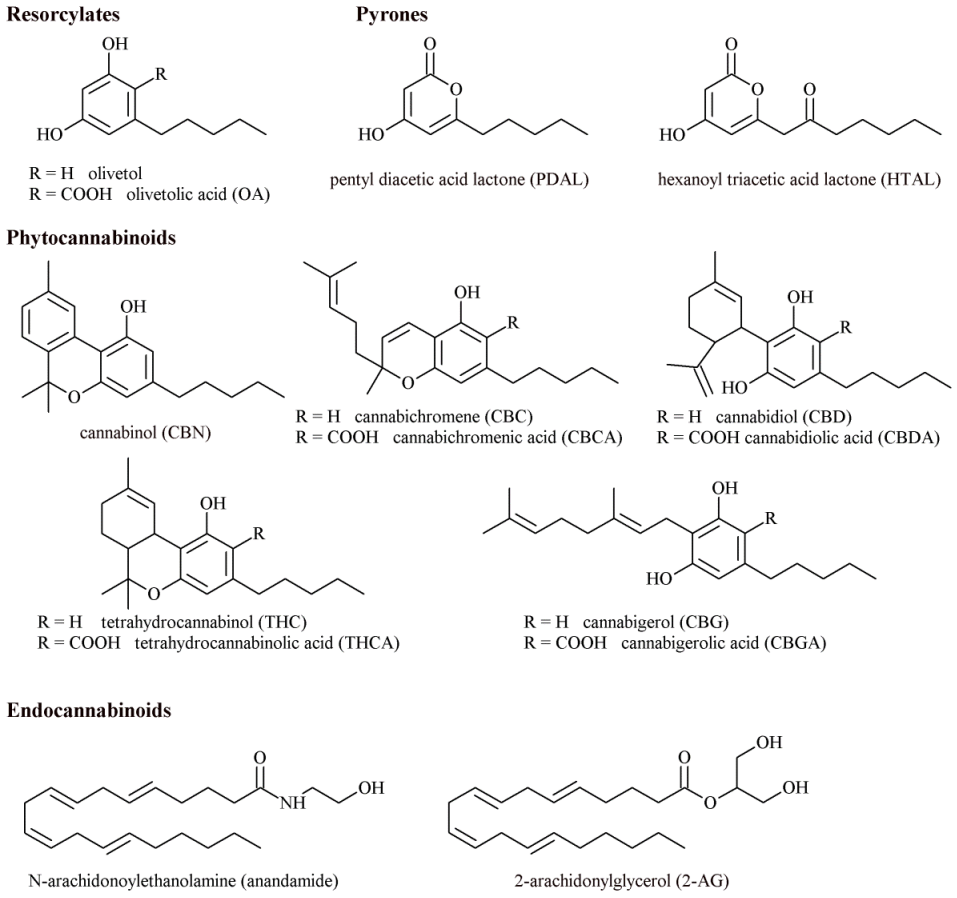


Figure 2.1 Structures for olivetol, OA, PDAL, HTAL and select cannabinoids

Resorcylates and pyrones are the main focus of this study; they are *in vitro* products from enzymes involved in the biosynthesis of phytocannabinoids. Phytocannabinoids are produced in plants, whereas endocannabinoids occur in animals.

Figure 2.2 A female cannabis plant.
Cannabis trichomes used in this study were obtained from plants such as this *Cannabis sativa* cultivar Finola. This plant has reached maturity evident by the developed inflorescence.

Some confusion also exists as to whether *Cannabis* is a monotypic or polytypic taxon; some authors believe only one taxon exists, *Cannabis sativa* L. (Small and Beckstead, 1973), whereas other authors claim that three distinct species exist: *C. sativa, Cannabis indica,* and *Cannabis ruderalis* (Schultes et al., 1974). This work will consider cannabis as an inclusive

species called *Cannabis sativa* and will not discriminate between sub-species and/or cultivars. Perhaps more importantly is the classification of the different chemical phenotypes. Varieties of cannabis grown as industrial hemp for the production of fibre and seeds generally contain lower amounts of tetrahydrocannabinolic acid (THCA), which is the precursor to the psychoactive THC, and higher amounts of the non-psychoactive compound cannabidiolic acid (CBDA). Drug-type cannabis, which is grown both for recreational and medicinal use, contains significantly more THCA than CBDA (Fetterman et al., 1971; Small and Beckstead, 1973; Toffoli et al., 1968).

Cannabinoids are primarily biosynthesized in specialized structures known as glandular trichomes (Kim and Mahlberg, 1997; Potter, 2009). Trichomes are differentiated epidermal outgrowths which are specialized in form and function in many plants (Schilmiller et al., 2008). Four types of trichomes are found on cannabis plants; non-glandular hair-type trichomes, and three secretory-type trichomes: bulbous, capitate-sessile, and capitate-stalked glandular trichomes (Dayanandan and Kaufman, 1976; Hammond and Mahlberg, 1973) (Figure 2.3). Non-glandular trichomes are non-secretory, unicellular, and are prominently cystolith hairs containing high concentrations of silicon. Amongst the glandular trichomes, the capitate-stalked trichomes (150 to 500 µm) are the largest accumulators of cannabinoids, developing during the mature flowering stage. They are similar to capitate-sessile trichomes, but include a stalk in addition to a secretory disc and a subcuticular cavity. The secretory disc, which is a convex rosette of 8 – 13 cells, is the site of synthesis for fatty acids, terpenes, and cannabinoids which accumulate in the subcuticular cavity, giving the capitate gland a globular appearance (Dayanandan and Kaufman, 1976). Several studies have implicated disc cells in glandular trichomes to be responsible for the

production and secretion of secondary metabolites (Wagner, 1991). Sirikantaramas et al. (2005) found that the presence of mRNA for the enzyme THCA synthase was exclusively detected by RT-PCR in cannabis trichome disc cells, and furthermore, when heterologously expressed in tobacco, a THCA synthase-green fluorescent protein localized to the plant's trichomes (Figure 2.3).

The glandular trichome is understood as being the site of unique gene expression related to the production of secondary metabolites. Some evidence was provided by Keene and Wagner who purified glands from tobacco that were capable of efficiently producing diterpenes and sucrose esters, although surrounding tissues (epidermal and subepidermal cells) could not (Keene and Wagner, 1985). The genomic analysis of glandular trichomes, which typically involves the construction of a cDNA library derived from isolated trichomes, has been successful in elucidating specialized metabolic pathways and identifying enzymes specific to the trichome (Fridman and Pichersky, 2005). The specialized peltate glands from sweet basil were successfully analyzed this way, leading to the identification of a methyltransferase responsible for the biosynthesis of methylchavicol (Gang et al., 2001; Iijima et al., 2004). Sesquiterpene biosynthesis in wild tomato *Solanum habrochaites* was also determined by evaluating trichome-specific EST data, leading to the identification of two terpene synthases (Sallaud et al., 2009). The lupulin glands from *Humulus lupulus* were also separated and used to create a large set of ESTs which were then used to identify an *O*-methyltransferase (OMT-1) required in the conversion of desmethylxanthohumol to xanthohumol (Nagel et al., 2008). Trichome-specific cDNA libraries have also been used for gene discovery in cannabis, resulting in the detection of

a type III PKS capable of synthesizing olivetol and an acyl-activating enzyme responsible for the generation of hexanoyl-CoA (Marks et al., 2009; Stout et al., 2012).

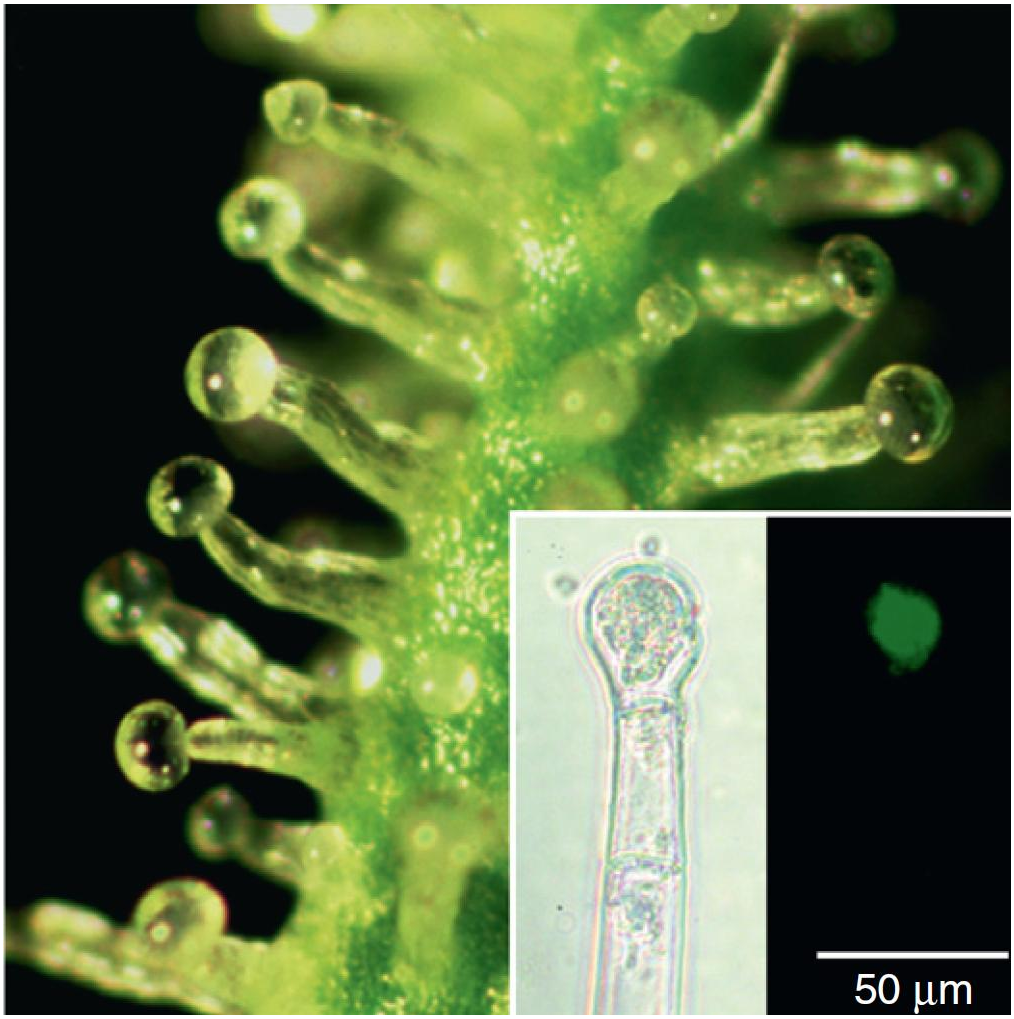


Figure 2.3 Trichomes are the site of cannabinoid biosynthesis.

The cannabinoids of *Cannabis sativa* are synthesized and stored in glandular trichomes found on the female flower. The inset shows the localization (green) of an enzyme of the cannabinoid biosynthetic pathway, tetrahydrocannabinolic acid synthase, using a green fluorescent protein-labeled probe to the head of a glandular trichome. Reproduced with permission from Jirschitzka et al. (2013).

2.1.2 Bioactivity of Cannabinoids

One of the earliest evaluations of cannabis in western medicine was conducted by William O'Shaughnessy in 1838, wherein he describes the anaesthetic, orexigenic, antiemetic, and anticonvulsant properties of the drug (O'Shaughnessy, 1843). Cannabinoids are now recognized as therapeutic agents for a number of debilitating ailments including chemotherapy-associated nausea, epilepsy, glaucoma, Parkinson's disease, dystonia, multiple sclerosis, and pain (Amar, 2006). The antineoplastic properties of certain cannabinoids, namely THC, have been reported as early as 1975, and although still requiring conclusive study, cannabinoids show great potential as antiproliferative, tumor-selective apoptotic agents (Brown et al., 2012; Munson et al., 1975). Generally, the mechanism of cannabinoids are related to their binding to the endogenous CB1 and CB2 receptors, but has also been shown to involve the ankyrin, melastatin, and vanilloid transient potential receptors (Pertwee et al., 2010). Cannabinoids acting on the classical cannabinoid receptors, which are G-protein coupled receptors, inhibit adenylcyclase and result in the modulation of the MAPK/ERK pathway.

2.1.3 The Current Understanding of Cannabinoid Biosynthesis

In the last 80 years several groups have proposed different cannabinoid biosynthesis pathways. The earliest proposal was offered by Simonsen and Todd (1942), after Adams (1940) showed that CBD could be converted to THC and CBN *in vitro*; they proposed that a cannabidiol-like compound formed by the condensation of menthatriene (2-methyl-5-prop-1-en-2-ylcyclohexa-1,3-diene) and olivetol cyclized to THC and later to CBN(Adams, 1940; Simonsen and Todd, 1942). Following the Birch (1957) proposal for phenol biosynthesis, Farmilo and Davis (1961) suggested that hexanoic acid condenses with three molecules of acetic acid yielding a

hexenedionic acid that could enolize to OA. Farmilo and Davis (1961) were the first to depict cannabinoids in their natural acidic form. These earlier accounts for cannabinoid biosynthesis were presented without the knowledge of cannabigerolic acid (CBGA) and were based on the belief that THC production was a result of CBD isomerization. Gaoni and Mechoulam (1964) realized that cannabigerol (CBG) was a missing intermediate in the cannabinoid biosynthetic pathway and suggested that two alternative pathways existed; one pathway each dedicated to the biosynthesis of either the acid or neutral cannabinoids, but both these pathways were believed to require the condensation of geranyl diphosphate (GPP) with olivetol (or OA in the case of acidic forms) (Gaoni and Mechoulam, 1964). Shoyama et al. (1970) later proposed a model for the cannabinoid biosynthesis pathway accounting for alternate forms of cannabinoids, namely cannabichromenic acid (CBCA) and CBGA. Shoyama et al. (1975) later published data from experiments with ^{14}C-labelled malonate/mevalonate and ^{3}H-labelled geraniol, indicating that CBGA is formed by the condensation of geraniol and a polyketide derived from malonate. These findings have provided the basis for some of the current understanding of the cannabinoid biosynthetic pathway.

Cannabinoids are terpenophenolic compounds that originate from the prenylation of the polyketide-derived OA with GPP, converting it to the first cannabinoid in the cannabinoid biosynthesis pathway, CBGA (reviewed by (Taura et al., 2007) (Figure 2.4). A geranylpyrophosphate:olivetolate transferase (GOT) was proposed to prenylate OA after observing that CBGA is produced by cannabis cell extracts when provided with GPP and OA (Fellermeier and Zenk, 1998). Importantly, this study concluded that OA served as the phenolic substrate for GOT and not its decarboxylated derivative, olivetol. The Zenk group provided

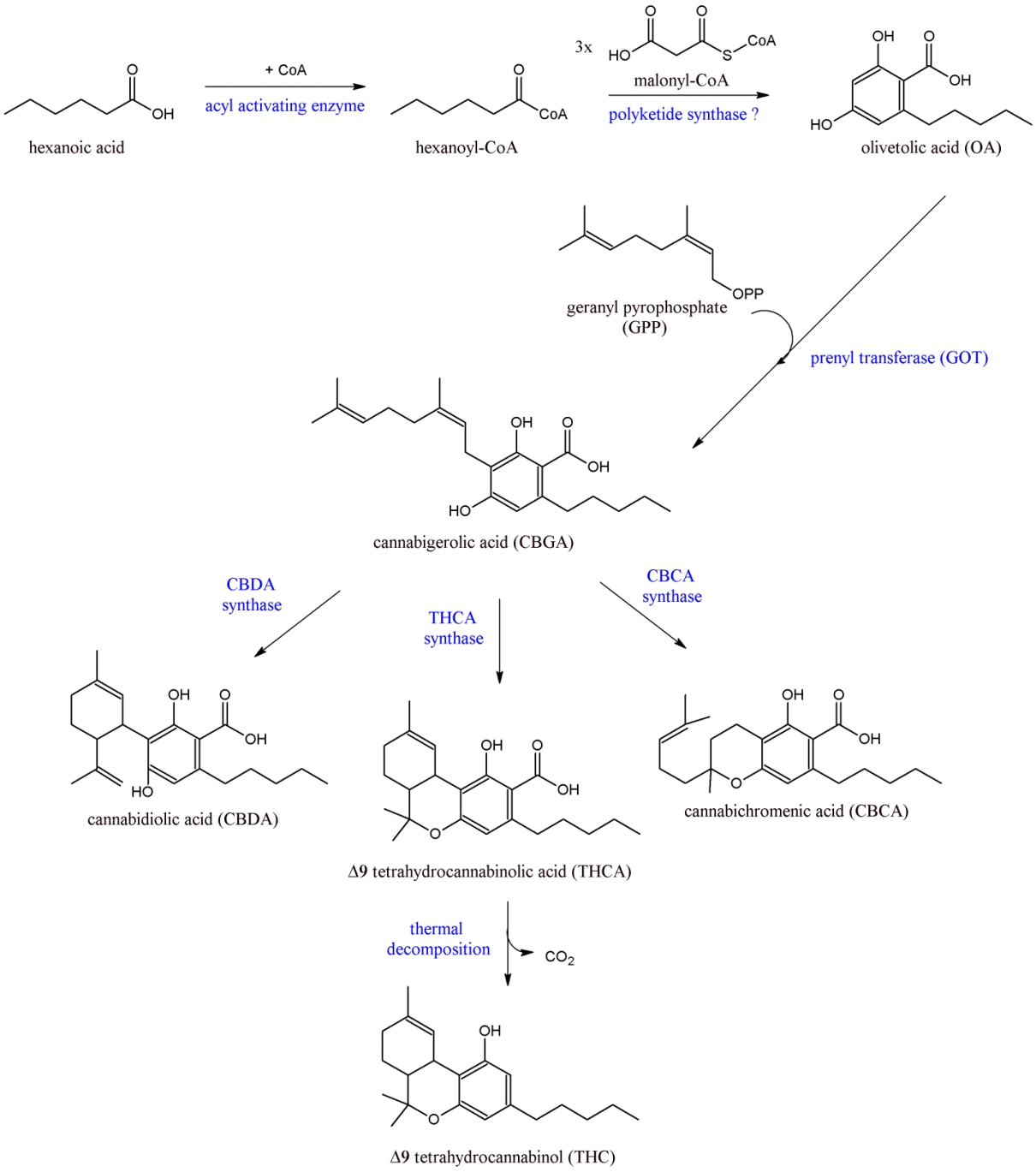


Figure 2.4 Understanding of cannabinoid biosynthesis based on current literature.

The intermediate CBGA is produced by the prenylation of OA by GOT. Whereas GPP is derived from the MEP pathway, OA has been proposed to be synthesized by a PKS capable of accepting hexanoyl-CoA esters once they are activated by an acyl-CoA synthetase. CBGA is converted to CBDA, THCA, CBCA by CBDA synthase, THCA synthase, and CBCA synthase, respectively.

evidence that the terpenoid substrates responsible for the generation of GPP, isopentenyl diphosphate (IPP) and dimethylallyldiphosphate (DMAPP), are derived from the deoxyxylulose phosphate pathway (Fellermeier et al., 2001). Their work strongly suggested that the phenolic moiety is a polyketide-derived compound.

The prenylation reaction resulting in the biosynthesis of CBGA is followed by the conversion into CBDA or THCA catalyzed by the enzymes, CBDA synthase and THCA synthase, respectively (Taura et al., 1996; Taura et al., 1995). It is important to note that the neutral form of CBGA, CBG is not accepted by THCA synthase or CBDA synthase as a substrate. CBGA also acts as the substrate for other cannabinoid synthases such as CBCA synthase (Morimoto et al., 1997; Morimoto et al., 1998) (Figure 2.4). These cannabinoids, their analogs, and a variety of others can be further classified into ten distinct classes of cannabinoids based on their structural backbone (Sanchez and Verpoorte, 2008).

2.2 Olivetolic Acid Biosynthesis

The concept that OA originates from a polyketide is based on the biosynthetic speculations made by both Collie and Birch, who independently proposed that phenolic natural products are derived from β-polyketones (Hill and Staunton, 2010). In this Section, an overview of type III PKSs and the early accounts of experiments suggesting the involvement of a PKS during OA biosynthesis is presented. Past attempts to detect and isolate a PKS capable of OA biosynthesis are reviewed.

2.2.1 Type III PKSs: Diversity, Structure and Mechanism

PKSs are enzymes that are distributed in the bacterial, fungal, animal and plant kingdoms. They are responsible for the generation of acetate-based β-polyketo compounds known as polyketides and are involved in the biosynthesis of many secondary metabolites. Many polyketides are

bioactive and play an important role in medicine, where they are used as antibiotics, immunosuppressants, statins, and cytostatics (Shen, 2003; Staunton and Weissman, 2001; Weissman, 2009). PKSs conduct the repeated condensation of malonyl esters with an appropriate starter molecule, resulting in an elongated polyketide, a process reminiscent of fatty acid biosynthesis. Not surprisingly, these enzymes are thought to have evolved from fatty acid synthases, which in addition to their similar activity share a highly conserved structural fold (Austin and Noel, 2003; Jenke-Kodama et al., 2005). PKSs are categorized as type I, II, or III: type I PKSs are large modular enzymes found in fungi, bacteria, and animals (Castoe et al., 2007; Keatinge-Clay, 2012); type II PKSs are part of multi-enzyme aggregates found in bacteria (Zhan, 2009); type III PKSs are small homodimeric proteins found in plants, bacteria, and more recently in fungi (Austin and Noel, 2003; Muggia and Grube, 2010).

Type III PKSs constitute their own family of enzymes that are believed to have evolved from fatty acid synthases, although they are structurally different than type I and II PKSs. The type III PKS family, also known as the chalcone/stilbene synthase family, is ubiquitous in higher plants and is involved in the biosynthesis of chalcones and stilbenes, which are themselves precursors for more complex secondary metabolites, including flavonoids and anthocyanins (Austin and Noel, 2003). Whereas type I and II PKSs require an acyl carrier protein (ACP) to activate malonyl extender units, type III PKSs utilize malonyl-CoA esters. Also, type III PKSs have greater diversity in their use of starter substrates compared to type I and type II PKSs, as is the case in the chalcone-forming reaction (Figure 2.5). The chalcone synthase (CHS) reaction performed by type III PKSs iteratively adds acetate units to a *p*-coumaroyl-CoA starter by the

decarboxylative condensation of malonyl-CoA. The resultant tetraketide intermediate is transformed into naringenin chalcone by an intramolecular Claisen condensation.

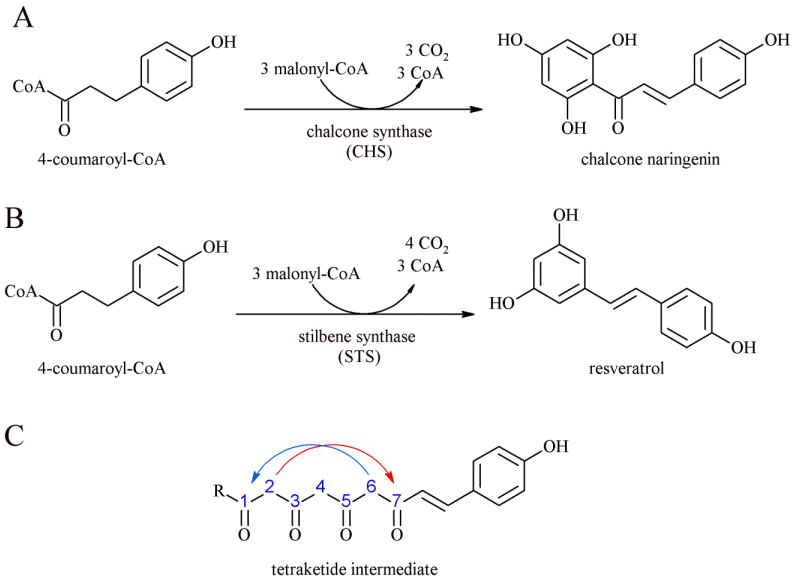


Figure 2.5 Type III PKS reactions.

Reactions include (A) the biosynthesis of chalcones such as chalcone naringenin by CHSs, and (B) the biosynthesis of stilbenes such as resveratrol by STSs. Both CHSs and STSs synthesize the same tetraketide intermediate, although (C) the cyclization of the reactive intermediate proceeds in either (blue arrow) a Claisen condensation or (red arrow) an aldol reaction. The R substituent indicates attachement to an enzyme's active site.

Type III PKSs do not solely synthesize naringenin chalcone, neither do they all involve a terminating Claisen condensation. Stilbene synthases (STS) are also type III PKSs which catalyze the synthesis of stilbenes, including the well known compound resveratrol (Figure 2.5) (Hathway and Seakin, 1959; Schröder et al., 1988; Takaoka, 1939). The initial steps of PKS reaction are identical in CHS and STS, where both form a tetraketide intermediate, although a different cyclization reaction takes place in the final steps of product formation. The CHS reaction depends on a $C_6 \rightarrow C_1$ Claisen condensation, whereas STSs catalyze a $C_2 \rightarrow C_7$ aldol reaction followed by a C_1 decarboxylation (Austin and Noel, 2003). The cyclization appears to

be controlled by electronic effects of specific AAs in respective enzymes. This was shown by Austin et al. (2004) after converting the CHS from *Medicago sativa* into a functional STS, which they engineered using *Pinus sylvestris* STS as a model. A third alternative reaction also conducted by these enzymes is a $C_5O \rightarrow C_1$ cyclization resulting in the generation of pyrone products (Figure 2.6). Pyrone products such as CTAL are commonly referred to as derailment products, and are regarded as incomplete reaction products. Austin et al. (2004) suggested that pyrone products may slowly and spontaneously cyclize when tetraketide intermediates are released by the enzyme, or that an unidentified enzyme may catalyze the final cyclization of released intermediates (Austin et al., 2004). It is important to note that some plants, such as *Gerbera hybrida* and *Piper methysticum*, do accumulate pyrone products (Eckermann et al., 1998; Seigler, 1998).

Stilbene carboxylate synthases (STCS) are a hypothetical subset of type III PKSs which possess a similar activity to STSs, but differ in that they retain the C_1 carboxyl moiety following a $C_2 \rightarrow C_7$ cyclization (Figure 2.6) Although there is less support for the existence of STCSs, they are assumed to biosynthesize phenolic acid metabolites *in planta*. One STCS, HmSTCS1a, has been reported to occur in *Hydrangea macrophylla*, an ornamental that produces the rare hydrangeic and lunularic acids (Figure 2.7). HmSTCS1a was isolated from *Hydrangea* cDNA and shares 65-70% sequence identity with other chalcone and stilbene synthases; however, the enzyme contains unusual residues in positions conserved in other CHS-like proteins. Although Eckermann et al. (2003) cloned HmSTCS1a, the hypothesized final product, 5-hydroxy-hydrangeic acid was never detected during assays, and instead the derailment product CTAL accumulated (Figure 2.7). The formation of the CTAL pyrone was regarded as an artefact

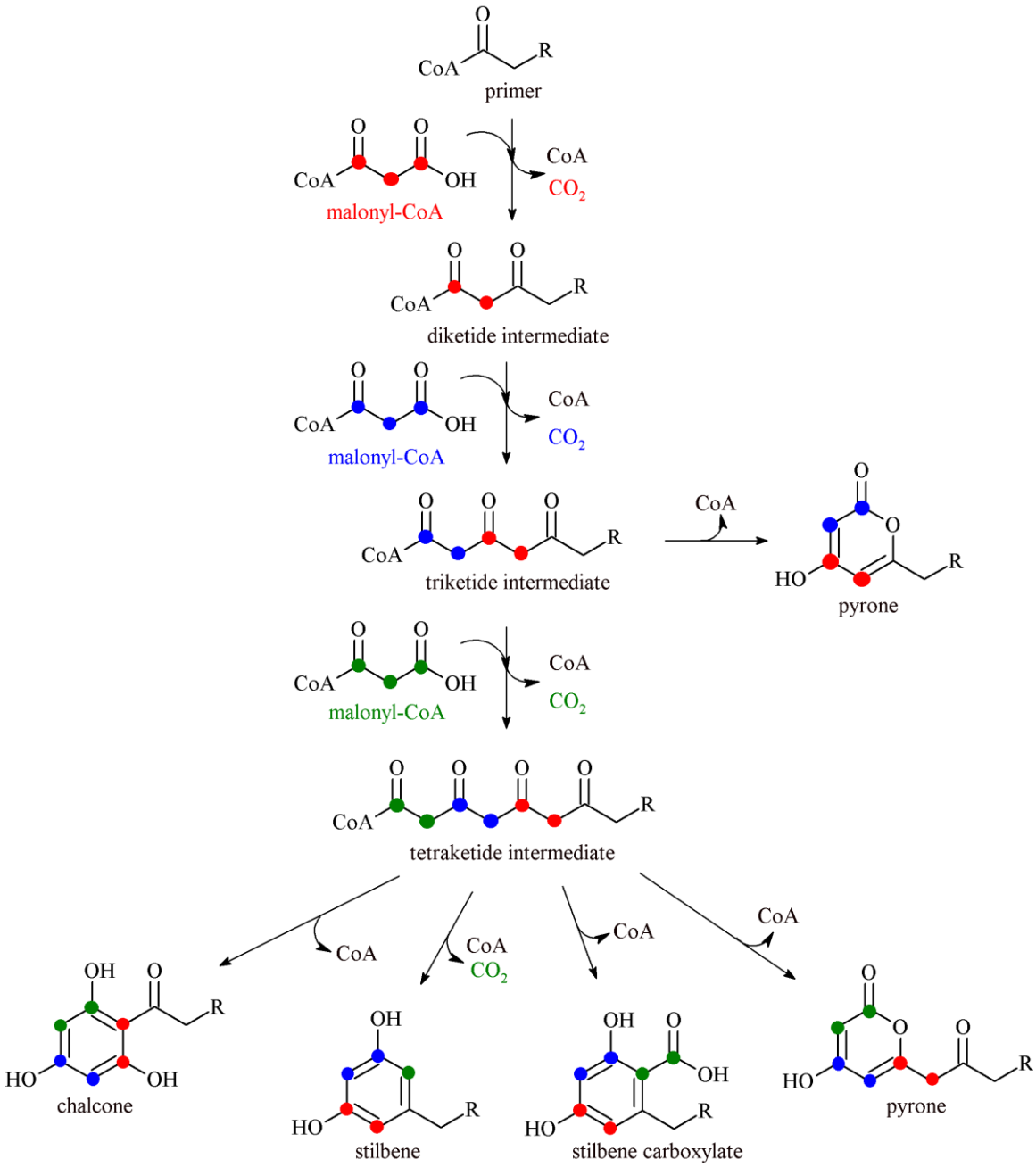


Figure 2.6 Polyketides are composed of acetate units.

Polyketide synthases use an array of different starter molecules (R) and iteratively couple them to acetate residues by the decarboxylative condensation of malonyl-CoA. The cyclizations of longer polyketide intermediates result in the formation of pyrones, whereas the cyclization of tetraketides can result in pyrones, chalcones, stilbenes, and stilbene carboxylates. Figure adapted from Schröder (2009).

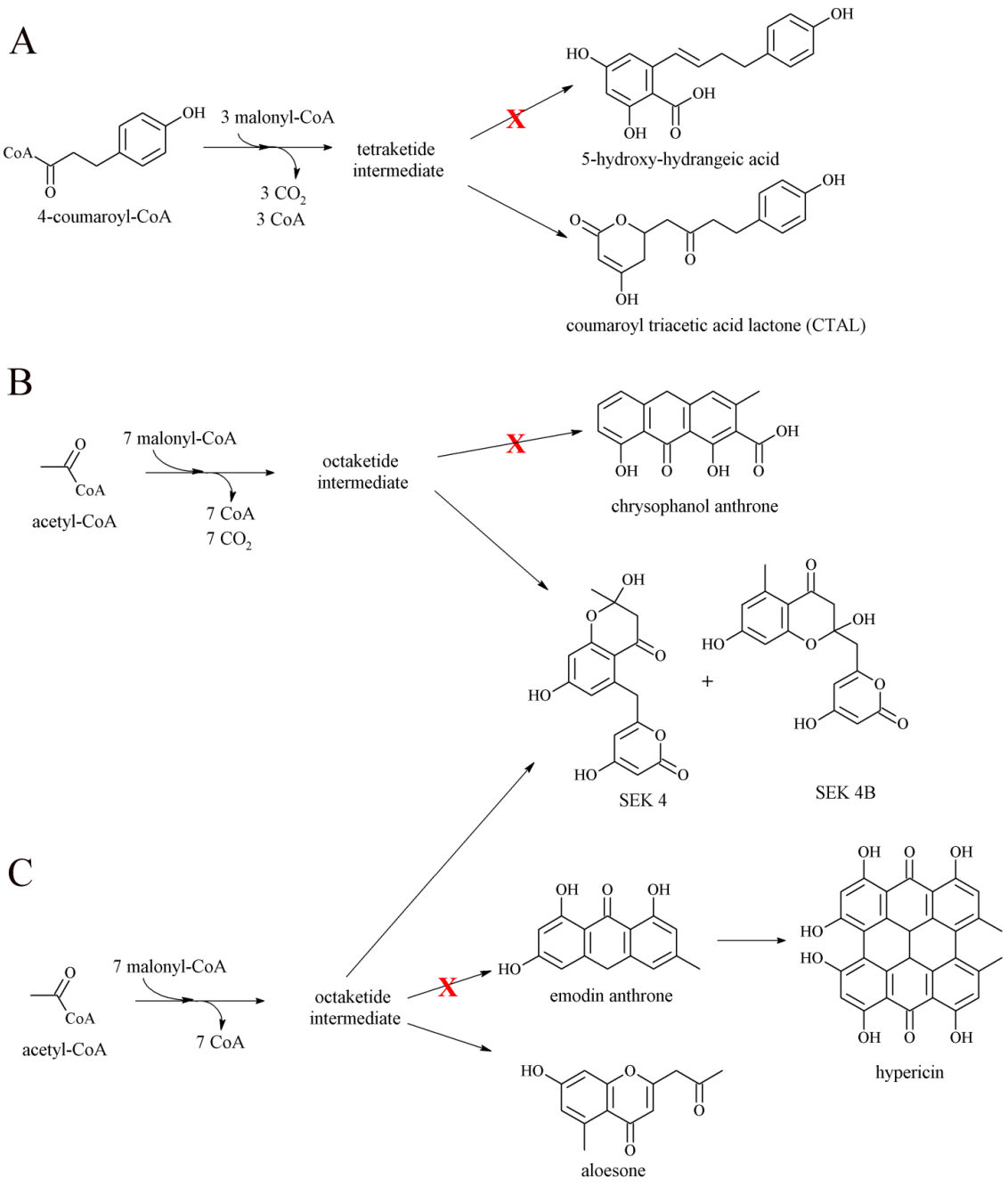


Figure 2.7 Type III PKS reactions can yield unanticipated pyrones.

Certain recombinant CHS-like PKSs putatively involved in the biosynthesis of natural products do not always yield the anticipated products when assayed *in vitro*. Examples include (A) the biosynthesis of 5-hydroxy-hydrangeic acid by *H. macrophylla* HmSTCS1, (B) the biosynthesis of chrysophanol anthrone by *A. arborescens* OKS, and (C) the biosynthesis of emodin anthrone by *H. perforatum* HpPKS2. Instead these reactions synthesized pyrone products.

of sub-optimal *in vitro* conditions that did not allow for the appropriate cyclization to occur. Eckermann et al. (2003) also described the *in vitro* biosynthesis of 5-hydroxylunularic acid, but only after a different starter substrate was used (*i.e.* dihydro-4-coumaroyl-CoA instead of 4-coumaroyl-CoA) and after the permethylation of assay products was performed.

Other examples of CHS-like enzymes incapable of cyclizing polyketide intermediates to the anticipated final products have also been reported in other plants. *Aloe arborescens* octaketide synthase is a type III PKS thought to be involved in the biosynthesis of chrysophanol anthrone, but which instead accumulates the pyrones SEK4 and SEK4b when the recombinant enzyme is assayed (Figure 2.7) (Abe et al., 2005). Similarly, HpPKS1 and HpPKS2 are two CHS-like enzymes putatively involved in the biosynthesis of the biologically active hyperforin and hypericin in *Hypericum perforatum* L. (St. John's Wort) (Karppinen and Hohtola, 2008). Although the production of emodin anthrone was anticipated, the *in vitro* analysis of the HpPKS2 enzyme produced tri- to octaketide products, as well as the octaketide pyrones SEK4 and SEK4b, and the heptaketide aloesone, yet no emodin anthrone was detected (Figure 2.7). Whether these enzymes are involved in the respective reactions, or another putative enzyme or co-factor is required for the proper cyclization of the polyketide chain is not known.

2.2.2 Instances of Olivetolic Acid Biosynthesis

The occurrence of OA is rare in nature, but has been reported to accumulate in the fungal, plant, and animal kingdoms. OA is found in *Primula obconica*, where it is used as a substrate for the natural products primin and miconidin (Horper and Marner, 1995). In animals, the release of OA from metapleural glands in some *Crematogaster* ants has been reported (Jones et al., 2005). OA is also recognized as a component of depsides in numerous lichens (Armaleo et al., 2011; Elix, 1974; Elix, 1994). Additionally, OA analogs that incorporate longer alkyl chains than the pentyl chain in OA, also occur in nature, and could presumably be the result of similar cyclization reactions since the carboxylate moiety is retained in the final products. These analogs are found in numerous fungi and plants including *Evernia prunastri* (oakmoss), *Merulius incarnates* (mushroom), *Grevillea robusta* (silver oak), and several cereals and grains (Baerson et al., 2010; Hassani and Razzoul, 2005; Jin and Zjawiony, 2006; Yamashita et al., 2008; Yamashita et al., 2010; Yao, 2011). More recently the first example of a type III PKS capable of forming alkylresorcylic acids was reported in *Oryza sativa* (rice) (Matsuzawa et al., 2010).

2.2.3 The Biosynthesis of Olivetolic Acid Requires a Polyketide Synthase

Most of the data suggesting that OA is synthesized via an acetate polyketide pathway was provided by heavy isotope labeling experiments conducted in cannabis. The earliest account for PKS involvement was provided by Shoyama et al. (1975), after ^{14}C-malonate was incorporated into cannabinoids during isotope feeding experiments (Shoyama et al., 1975). The involvement of a PKS was also implied after Fellermeier et al. (2001) used ^{13}C-glucose to discern the positions of the isotopes incorporated into cannabinoids. Similarly, after conducting ^{3}H-acetate labeling studies in *Primula obconica*, Horper et al. (1996) suggested that a PKS was involved in

the biosynthesis of OA, as well as predicting that the starter moiety was hexanoic acid. Results from these groups indicate that an acetate-derived polyketide corresponds well to the NMR labelling patterns they obtained. By the year 2000, the search for an OA-producing enzyme had begun, with the focus that a type III PKS was responsible for the generation of the necessary substrate.

2.2.4 Type III PKSs from Cannabis

One of the earliest attempts to detect polyketide synthase activity in cannabis was conducted by assaying the crude extracts from various parts of the plant (Raharjo et al., 2004a). The extracts were incubated with hexanoyl-CoA and malonyl-CoA esters, and although the authors were able to detect olivetol, they found no OA. These experiments provided the basis for further rigorous examination into the enzyme(s) responsible for OA biosynthesis. Raharjo et al. (2004c) continued their search using a proteomics approach that compared the differential expression of proteins and mRNA within tissues capable and incapable of cannabinoid biosynthesis. The authors were unable to determine any of the cannabinoid biosynthetic enzymes during their proteomic analysis of leaves, flowers, and trichomes. Furthermore, Raharjo et al. (2004c) also attempted to identify cannabis PKSs using a *Pinus sylvestris* chalcone synthase antibody, and were partially successful in that they detected a PKS similar to the *H. lupulus* chalcone and phloroisovalerophenone synthases (Matousek et al., 2005; Paniego et al., 1999). Determined to isolate the recently detected PKS, Raharjo et al. (2004b) conducted degenerate reverse transcriptase-polymerase chain reaction (RT-PCR) using primers modeled after the *H. lupulus* CHS and phloroisovalerophenone synthase, and did successfully clone the first cannabis PKS. This enzyme (deposited in Genbank AAL92879) showed chalcone and phloroisovalerophenone

synthase activity, but was incapable of olivetol and OA biosynthesis, and was assigned as simply a cannabis CHS.

In 2009, a new cannabis PKS named olivetol synthase (OLS; annotated BAG14339) was isolated from cannabis leaves using degenerate primers, and reported to biosynthesize olivetol from hexanoyl-CoA and malonyl-CoA (Taguchi et al., 2008; Taura et al., 2009). OLS also synthesizes two other products, namely a triketide pyrone called pentyl diacetic acid lactone (PDAL) and a tetraketide pyrone named hexanoyl triacetic acid lactone (HTAL) (Figure 2.8).

Almost simultaneously with the Taura et al. (2009) publication, Marks et al. (2009) also reported the identification of a type III PKS possessing olivetol synthase activity. This discovery was made after a trichome-specific cDNA library was generated in an attempt to identify cannabinoid biosynthetic genes (Marks et al., 2009). Marks et al. (2009) reported a cDNA library of 1075 unigenes, of which three were putative type III PKSs named CAN24, CAN383, and CAN1069. Whereas CAN1069 was identical to the cannabis CHS previously reported by Raharjo et al. (2004b), CAN24 and CAN383 were reported as novel PKSs due to their divergent sequence. The authors characterized CAN24 and speculated that it produced pyrones and olivetol, since their results were inconclusive due to unsatisfactory HPLC analysis. It is now clear that CAN24 is identical to OLS based on sequence comparisons. Despite the successful identification of the novel type III PKS in cannabis, the characterization of OLS/CAN24 indicated that OA was not synthesized, and instead only its decarboxylated analog, olivetol, could be produced.

Another noteworthy analysis of cannabis PKSs was reported by Flores-Sanchez et al. (2010) after isolating a variety of PKS cDNAs from cannabis glandular trichomes by RT-PCR.

The authors described five different PKSs, PKSG1 (annotated EU551163), PKSG2 (annotated EU551164), PKSF3 (annotated EU551162), PKSG4 (annotated EU551165), and PKSG5 (annotated EU551166), that are 97% homologous over their 385 amino acid span. The largest difference between these PKSs appears in the size of their untranslated regions, which Flores-Sanchez et al. (2010) attributed to alternative transcription initiation and polyadenylation sites. None of the PKSs identified were identical to OLS/CAN24, although they all shared greater than 97% sequence similarity at the amino acid level with it. This small difference in sequence is believed to be due to the varieties of cannabis used in each study, although these individual isozymes have not been fully characterized. Flores-Sanchez et al. (2010) noted that the biochemical characterization of these PKSs should be conducted to better understand their function and diversity.

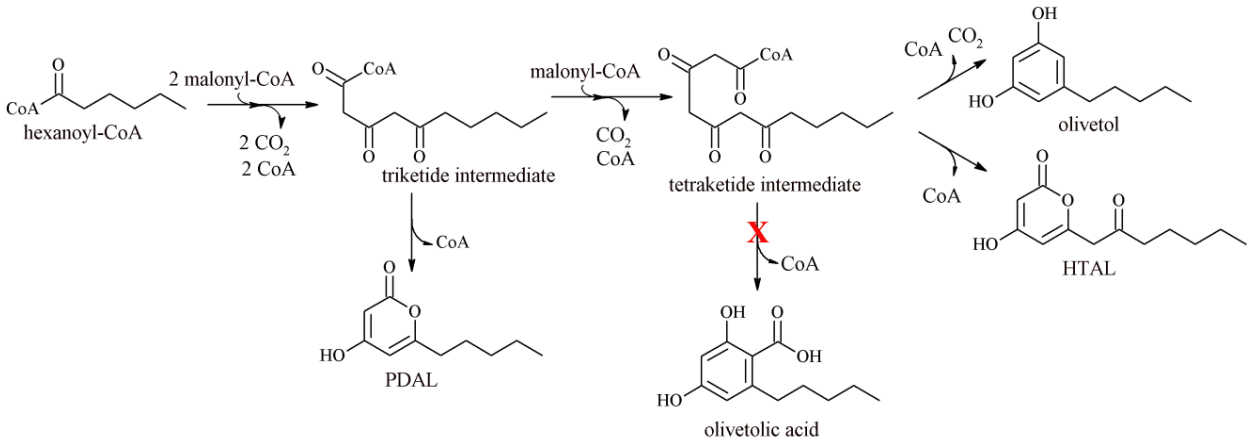


Figure 2.8 A PKS from cannabis is capable of producing PDAL, HTAL, and olivetol.
Although this type III PKS can produce olivetol, it cannot produce OA, the suspected intermediate required for further cannabinoid biosynthesis.

2.3 Polyketide Cyclization and DABB Proteins

2.3.1 Polyketide Cyclases

Generally, in all types of PKS systems, the biosynthesis of aromatic polyketides is a process that not only depends on the iterative condensation of acetate units, but also on the proper cyclization and tailoring of intermediates. A polyketide intermediate may be cyclized into several different products, for which only one specific enzyme may be capable of achieving. Cyclization reactions, with limited exception, differ substantially between fungal, bacterial, and plant systems, and have each been shown to employ different strategies to achieve the diverse product biosynthesis observed in nature. The biosynthesis of fungal polyketides is accomplished by megasynthases that incorporate a folding pattern (F mode) completely different than the folding pattern used by the actinobacterial species (S mode), which use dissociable multi-enzyme complexes during polyketide biosynthesis (Thomas, 2001). The S and F modes proposed by Thomas (2001) involve the cyclization of the initial ring in polyaromatic polyketides, and have been observed during many ^{13}C-acetate labeling experiments allowing the comparison of fungal and streptomycete polyketide biosynthesis, such as viridicatumtoxin biosynthesis in *Penicillium viridicatum* with oxytetracycline biosynthesis in *Streptomyces rimosus* (de Jesus et al., 1982; Thomas and Williams, 1983) (Figure 2.9). Based on these observations, Thomas (2001) designed a general set of guidelines for S and F modes, which indicated that F-mode cyclization results in the synthesis of an initial cyclohexane ring containing two intact ^{13}C-acetate moieties attached to two adjacent odd-numbered chains. Conversely, S-mode cyclization results in a cyclohexane containing three intact ^{13}C-acetate units with two even-numbered adjacent chains. Furthermore, the cyclization of polyketides is not simply a matter of S or F designation, but is also dependent

on several influencing factors including the oxidative state of the polyketide (especially around C9), the priming-compound used, the size of the substrate binding cavity or product template domain, the fusion of other polyketide intermediates, and the presence of other polyketide cyclases (Zhou et al., 2010). In their conceptual review, Zhou et al. (2010) have further categorized type II PKS cyclization based on the initial ring-folding pattern commonly occurring at the C7-C12 position, which generates the tetracycline, aureolic acid, anthracycline, and benzoisochromanequinone backbones, or at the C9-C14 position, which gives rise to the tetracenomycin scaffold. Zhou et al. (2010) suggest that the folding pattern mostly governed by the chain length factor, which does not only determine the size of a nascent polyketide, but also coordinates the polyketide after it buckles back, due to the limited space in the substrate-binding cavity. Additionally, these authors indicate that bacterial cyclases directly influence the initial ring cyclization pattern, such as seen with ZhuI and TcmN aromatases, which are able to coordinate C7-C12 and C9-C14 cyclizations, respectively. Although ZhuI and TcmN share structural similarity, a comparison of their structures indicates that ZhuI contains a smaller substrate-binding pocket which forces the nascent polyketide to bend at C7 instead of C9 (Ames et al., 2011). Ketoreductases, such as *Streptomyces coelicolor* ActIII, are involved in the reduction of polyketides at the C9 position, and have also been shown to influence C7-C12 cyclization in systems incapable of such cyclizations, such as in *Gibberella fujikuroi* PKS4 (Hadfield et al., 2004; Ma et al., 2007).

In plants, the folding of polyaromatic polyketides is accomplished by the simultaneous use of both S and F modes (Thomas, 2001). Aside from the CHS/STS folding pattern previously discussed (Section 2.2.4), Thomas (2001) refers to 2-methylanthraquinone chrysophanol and its

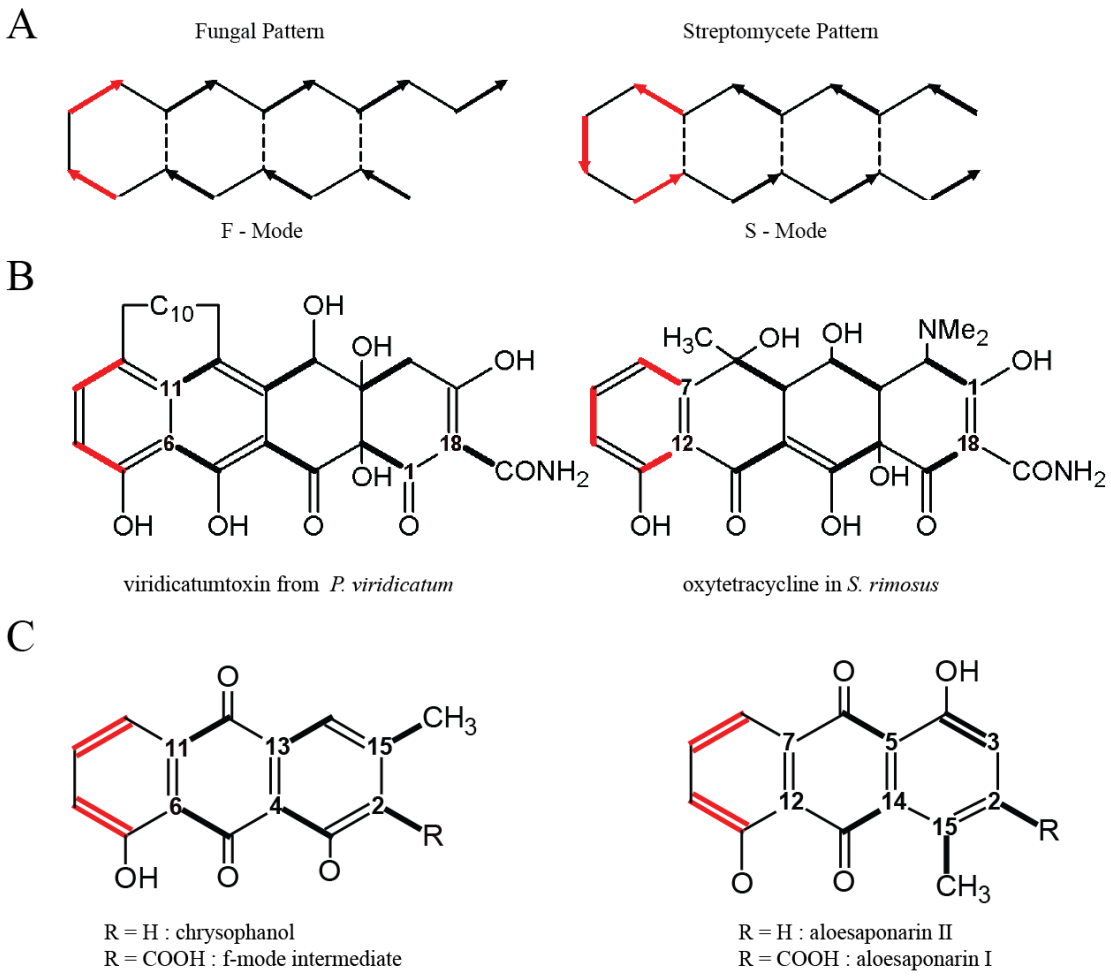


Figure 2.9 The F and S modes of polyketide cyclization.

(A) Fungal and streptomycete folding patterns differ during initial ring cyclization, where in F mode a C6-C11 bond forms, and in S mode a C7-C12 bond is formed. (B) The incorporation of ^{13}C-acetate during the biosynthesis of viridicatumtoxin and oxytetracycline is one example that illustrates F and S modes, respectively. (C) It has been suggested that the biosynthesis of chrysophanol and aloesponarin in *Aloe* occurs via a bimodal folding process. Chrysophanol can also be biosynthesized by *P. islandicum*, whereas aloesaponarin can be produced in streptomycetes. Figure adapted from Thomas (2001).

isomer, 1-methylanthraquinone aloesaponarin II, to exemplify the bimodal process in plants (Figure 2.9). Whereas both compounds are produced in *Aloe* species, chrysophanol is also produced in *Penicillium islandicum* through a decarboxylative F-mode cyclization, and aloesaponarin I can be biosynthesized in recombinant streptomycetes via an S-mode folding pattern. The occurrences of both these compounds in *Aloe* suggest an F/S bimodal cyclization process is responsible.

2.3.1.1 Secondary Polyketide Cyclases

Following primary ring cyclization, type II PKS systems use dissociable cyclases to fold the remaining rings during polyketide biosynthesis. These cyclases may act independently or may require other cyclases to synergistically achieve the correct folding pattern resulting in either linear or angular fused polyketides. In their review, Zhou et al. (2010) have divided these polyketides by their primary ring cyclization, but also by whether or not they form linear or angular fused rings, as well as by their secondary, tertiary, and in some cases quaternary cyclizations which result in several families of compounds that share a similar scaffold. For example, tetracyclines and aureolic acids are classified together as linear fused aromatics that are the result of C7–C12, C5–C14, C3–C16, and C1–C18 regiospecific cyclizations, whereas tetracenomycins are also linearly fused aromatics, but form their own class of compounds since they are generated following C9–C14, C7–C16, C5–C18, and C2–C19 cyclizations. The involvement of regioselective cyclases is required regardless of which class of fusion the polyketides belong to. In the case of tetracycline biosynthesis, it has been demonstrated that, in *S. coelicolor*, the cyclization of the four rings in the naphthacene backbone are achieved by only two cyclases; OxyK is responsible for the initial C7-C12 cyclization and OxyN is responsible for

the second ring C5-C14 cyclization (Petkovic et al., 1999; Zhang et al., 2007). The third ring is believed to result from spontaneous cyclization, but no enzyme has been identified as being responsible for the fourth ring cyclization. Similarly, the biosynthesis of tetracenomycin C requires a primary ring, and in this case secondary ring, cyclase, but also requires TcmI to complete the C2-C19 cyclization of the fourth ring. The mechanism for the third ring cyclization is still not well understood (Zhou et al. 2010).

2.3.1.2 A Possible Role for Polyketide Cyclases in Cannabinoid Biosynthesis

In an attempt to better understand the cyclization to OA, a general review of similar mechanisms incorporating the cyclization of polyketide intermediates was conducted. This overview of cyclization mechanisms capable of retaining a carboxylate moiety in the final or intermediate products led to the identification of similar phenomena occurring in Actinobacteria. Actinobacteria are a rich source of diverse secondary metabolites from which hundreds of antibiotics have been isolated. Particularly, the biosynthesis of tetracenomycin in *Streptomyces glaucescens* played an important role in the identification of enzymes putatively involved during the cyclization of OA. The biosynthesis of tetracenomycin is a multistep reaction that recruits a type II PKS responsible for the generation of a nonaketide intermediate (*via* TcmJKLM), that is initially cyclized to TcmF2 (*via* TcmN), and is cyclized once more, in a step that retains the carboxyl moiety by a type II polyketide cyclase known as TcmI (Thompson et al., 2004). The similarity of this reaction led to the exploration of similar proteins in cannabis, which eventually led to the successful progress described in this thesis. TcmI is classified as a polyketide cyclase (PKC), a family of proteins that harbor a specific dimeric $\alpha + \beta$ barrel (DABB) domain.

2.3.2 DABB Proteins

The DABB domain is a common structural motif found in many proteins distributed throughout nature that form a superfamily of structurally related proteins with some very common features. Some of these proteins play critical roles during important processes involving drug biosynthesis, environmental detoxification, and pathogenicity. Although DABB proteins share striking structural similarity, they are not necessarily derived from a common evolutionary origin and may not share strong sequence similarity. The dissimilarity in protein sequence is often seen at the family level where members sharing less than 20% sequence identity may have structures that superimpose to a root mean square deviation (rmsd) of $\leq$ 2.0Å. Proteins are typically categorized based on their similarities in sequence, function, and structure; traits that are classified differently by the available bioinformatic tools which lead to varying degrees of clustering by accessible databases (Martinez, 2011). DABB proteins are not excluded from this ambiguity and are represented differently depending on the database used. For example, the Structural Classification of Proteins database (SCOP) recognizes DABB proteins as a superfamily based on their three-dimensional domain, and partitions the superfamily into 23 sub-families (annotated 54909), whereas the PFAM database recognizes the signature motif for the DABB clan (annotated CL0032) which is further separated into 15 distinct families (Bateman et al., 2004; Murzin et al., 1995). As well, the two popular web-based resources, Interpro and the Conserved Domain Database (CDD), which similarly integrate information from several databases, distinguish the DABB superfamily very differently; Interpro recognizes DABB proteins as the parent of several sub-groups/families (annotated IPR011008), one of which the CDD narrowly defines as the DABB superfamily (Marchler-Bauer et al., 2011; Mulder et al.,

2005). In this study, the superfamily of DABB proteins is recognized as a larger superfamily represented by both SCOP and PFAM, although it focuses more on some functionally similar families, namely the antibiotic biosynthetic monooxygenases (ABMs), the PKCs, and the stress-responsive α + β barrel proteins (SRABBs).

2.3.2.1 The Structure and Occurrence of DABB Proteins

There are many protein structures available for DABB proteins that display common fundamental features. DABBs are α + β - classed structures displaying a repeated ferredoxin-like βαβ fold (Figure 2.10). These structurally analogous domains are characterized by independent segments of α-helices pressed against a 4-5 strand β-sheet, sometimes referred to as α-β sandwiches due to the distinct layering of the α-helices and β-sheet (Orengo and Thornton, 1993). DABB proteins typically dimerize at the β-sheet interface and form an anti-parallel β-barrel structure (Figure 2.10). The α-helices found in the repeated βαβ elements are sometimes segmented or kinked and form 1-2 arches which protrude away from the β-sheet, shaping a crevice between the secondary structures. This crevice is purported as the active site in several DABB members (Goblirsch et al., 2010; Sciara et al., 2003; Sugano et al., 2007). Although many DABB proteins display higher-ordered quaternary structure, they typically form dimers which may or may not continue to oligomerize. The formation of dimers is partly due to hydrophobic interactions between the residues of two subunit β-sheets that are not necessarily conserved yet result in similar β-barrel structures. Sciara et al. (2003) pointed out two highly conserved residues, Tyr63 and His52, in *Streptomyces coelicolor* ActVa-Orf6 homologs, which facilitate dimer stabilization through tyrosine-tyrosine aromatic stacking and tyrosine-histidine hydrogen bonds and demonstrated that replacement of His52 by site-directed mutagenesis (SDM) led to a

severe loss of activity due to the disruption of the dimer interface. Additionally, in ActVa-Orf6 each monomeric C-terminus was shown to contribute to dimer formation by swapping strands with the opposing β-sheet, also contributing to the bottom portion of the active site. But these features are not universal in DABB proteins. The *Staphylococcus aureus* IsdG, which shares only 15% identity at the amino acid level with ActVa-Orf6, yet can be superimposed to a rmsd of 1.92 Å, does not use the same residues as ActVa-Orf6 to achieve dimerization, and relies on different hydrophobic interactions to achieve its similar structure (Wu et al., 2005). Essentially, the involvement of hydrophobic residues is necessary for dimer interaction normally seen at the β-sheet interface. As well, strand swapping is not always observed in DABB members as it is in ActVa-Orf6. In *Escherichia coli* YgiN, there is an extended loop between the β2 and β3 strands which helps to form the opposing monomer's β-sheet instead of the strand swapping observed in ActVa-Orf6, yet structural similarity between ActVa-Orf6 and YgiN is maintained to an overall rmsd of 1.9 Å (Adams and Jia, 2005).

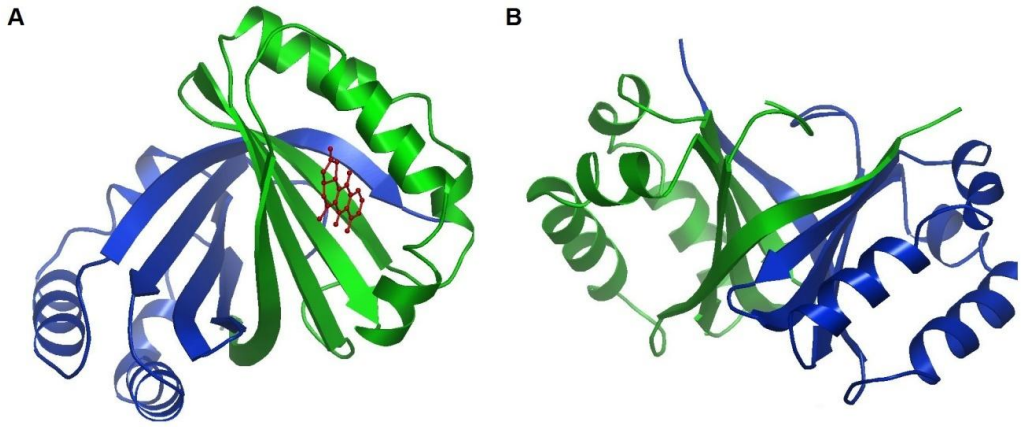


Figure 2.10 The structure of *S. coelicolor* ActVa-Orf6.

(A) ActVa-Orf6 is pictured as a dimer. The (blue) monomer carboxyl terminus is seen strand swapping with the opposing (green) monomer β-sheet strengthening the dimeric interaction and completing the β-barrel structure. The (red) substrate (Nanaomycin D) is shown in the putative active site found between the α-helices and β-pleats. (B) ActVa-Orf6 is shown at a different angle allowing the β-barrel structure to be seen. Figure adapted from Sciara et al. (2003).

2.3.2.2 The Function of DABB Proteins

DABB proteins are involved in catalyzing a diverse set of reactions including the oxygenation of polyaromatic compounds, isomerization of methyl muconolactone, and the cyclization of polyketides (Figure 2.11). The functionality of the different DABB proteins has been instrumental in the assignment of subfamilies within the superfamily of proteins. As the name implies, the ABM family is involved in the oxygenation of cyclic polyketide-derived antibiotic compounds. The *S. coelicolor* ActVa-Orf6 serves as the prototypical monooxygenase representative of the ABM family from which the ABM family name was initially derived. ActVa-Orf6 catalyzes the oxidation of 6-deoxydihydrokalafungin (6-DDHK) to dihydrokalafungin (DHK), an intermediate in the actinorhodin biosynthetic pathway (Kendrew et

al., 1997) (Figure 2.11). Other members of this family include *Streptomyces galilaeus* AknX and *Streptomyces nogalater* SnoaB involved in aclarubicin and nogalamycin biosynthesis respectively (Raty et al., 2002; Ylihonko et al., 1996).

Methylmuconolactone methyl isomerases (MLMIs) are involved in the degradation of methyl and chloro-substituted aromatics following *ortho*-(intradiol) cleavage. The cleavage of 4-methylcatechols form 4-methylmuconolactones which cannot be further processed by the 3-oxoadipate pathway unless they are initially isomerized to 3-methylmuconolactone by MLMIs (Figure 2.11). These enzymes have been described in *Cupriavidus necator* JMP132 (Pieper et al., 1990), *Rhodococcus rhodochronous* N75 (Bruce et al., 1989), and *Pseudomonas reinekei* MT1 (Camara et al., 2007). In addition to 4-methylmuconolactone isomerization, MLMIs have been shown to synthesize 3-methylmuconolactone using 1-methylbislactone as substrate indicative of MLMI's ability to open and close lactone rings (Pieper et al., 1990).

The PKCs form a separate family of DABB enzymes that are found exclusively in Actinobacteria. This family is exemplified by TcmI, the only functionally and structurally characterized PKC from this family to date. TcmI is part of the tetracenomycin biosynthesis gene cluster found in *S. glaucescens* and was shown to cyclize Tcm F2 to Tcm F1, both intermediates involved in the biosynthesis of the antitumor antibiotic Tcm C, without requiring a cofactor or metal ion (Figure 2.11) (Shen and Hutchinson, 1993a). It is important to note that this cyclization occurs while retaining the carboxyl moiety in the intermediates. Although the functionalities of these DABB proteins are quite diverse, the ability of the enzymes to bind and modify larger cyclic compounds appears universally conserved.

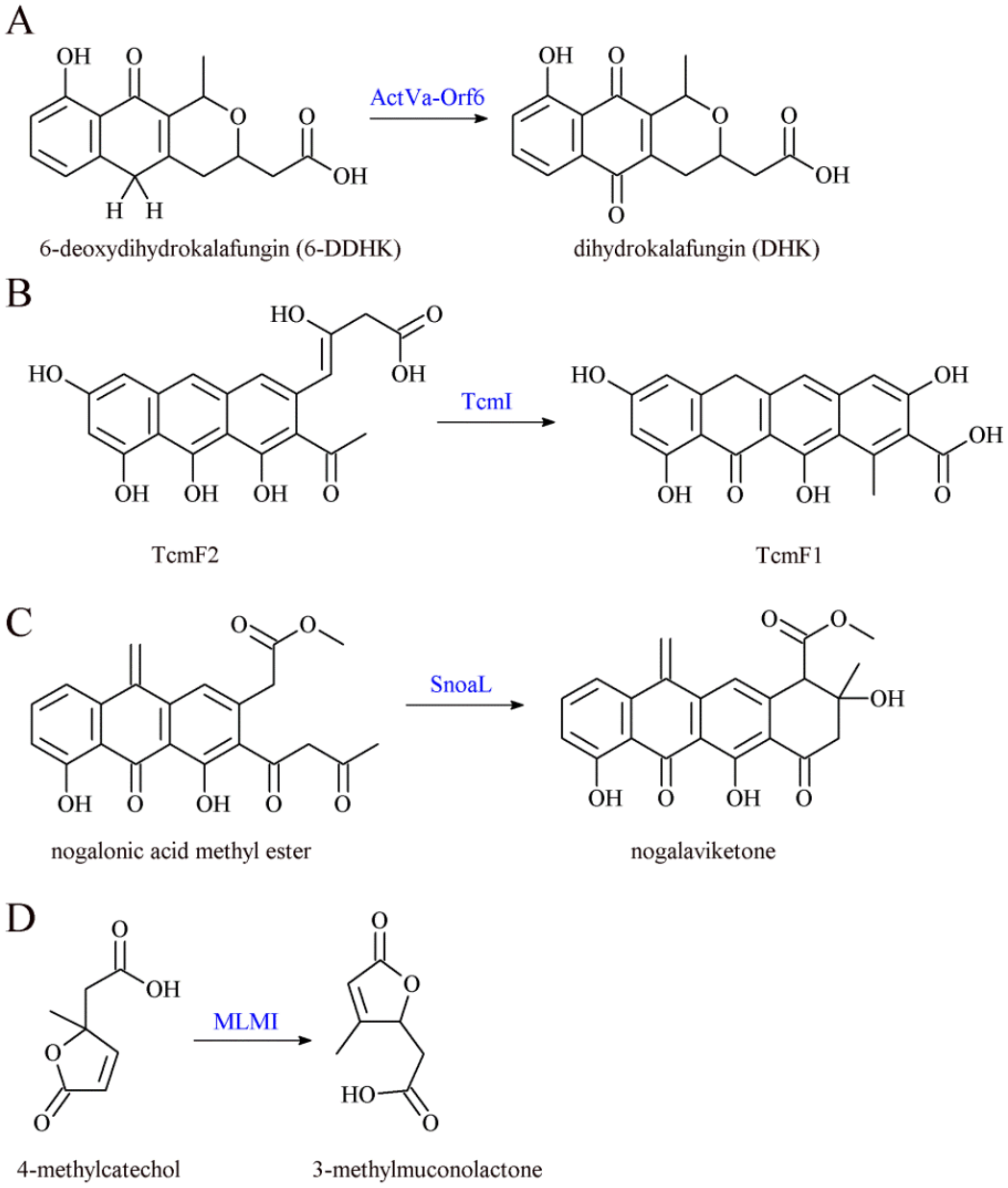


Figure 2.11 Examples of DABB protein activity.

(A) The oxidation of 6-DDHK to DHK is catalyzed by ActVa-Orf6. (B) TcmF2 is cyclized to TcmF1 by the TcmI cyclase. (C) Cyclization reactions which retain the carboxyl moietiy throughout are conducted by TcmI and are reminiscent of those coordinated by SnoaL enzymes which are sometimes fused to DABB motifs,. (D) MLMIs convert 4-methylcatechol to 3-methylmuconolactone.

2.3.3 Aldolases

Aldolases are enzymes capable of making or breaking an aldol bond, which are catoregized as being either class I or class II. Whereas class I aldolases, occurring primarily in animals and higher plants, form Schiff bases with their substrates *via* a lysine residue in the enzyme active site, the class II aldolases, which occur in bacteria and fungi, use a Zn^{2+} cofactor that acts as a Lewis acid during the reaction. The aldolases are important enzymes to consider in the context of this study, and are later discussed with regard to their distinct mechanisms; see Gijsen (1996) and Fessner (1996). Aldolase reactions are notable since they result in a donor retaining its carboxyl moiety, a phenomena which also occurs during the biosynthesis of OA.

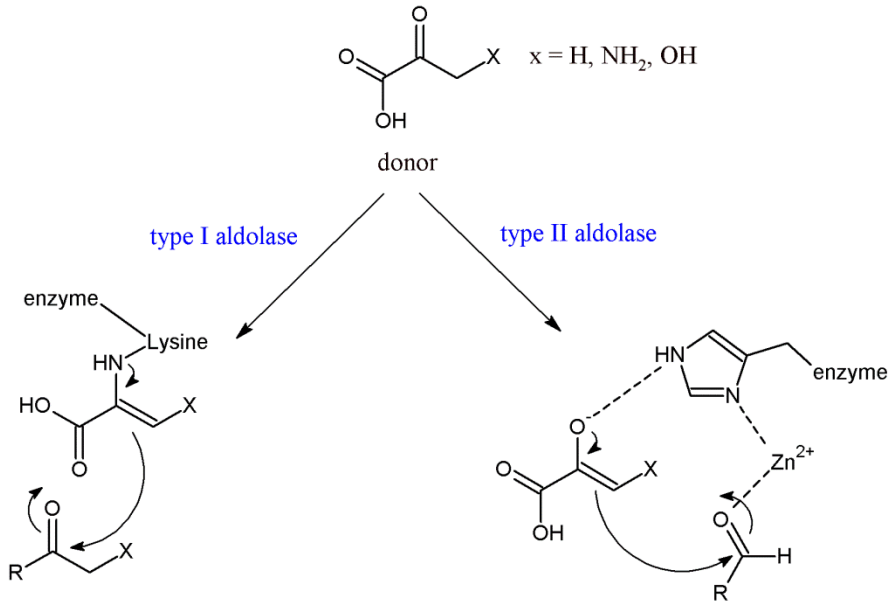


Figure 2.12 Type I and II Aldolase Reactions

In the case of type I aldolases, the donor compound is attached to the enzyme by a Schiff base prior to the aldol reaction, whereas type II aldolase reactions require a metal to coordinate the donor and the acceptor compounds. The substituent (R) shown in the acceptor compound may be a variety of monosaccaryl or oligosaccharyl chains.

3. THE TKS REACTION

3.1 Introduction

This study was preceded by experiments on trichome-specific cDNA library containing over 4,000 unigenes generated from high-THCA cannabis (Stout et al., 2012). This library was created by sequencing cDNA synthesized from glandular trichome mRNA and was used as a tool for the identification of enzymes involved in the biosynthesis of cannabinoids, one of which Stout et al. (2012) described as a hexanoyl-CoA synthetase. Displayed in Table 3.1 are the most highly expressed transcripts identified in the cDNA library along with their descriptors determined by BLAST searches against the Universal Protein Resource database (Uniprot). The most highly expressed transcript in this cDNA library was CL1Contig3, a sequence with 74% protein identity with *Humulus lupus* CHS (annotated BAB47196). The involvement of a type III PKS in the biosynthesis of OA had been previously suggested (Fellermeier et al., 2001; Horper and Marner, 1996; Shoyama et al., 1975), and soon after the identification of the type III PKS candidate by the Page laboratory (unpublished), several cannabis PKSs were reported in the literature (Flores-Sanchez et al., 2010; Marks et al., 2009; Taura et al., 2009). These reports included a similar trichome-derived cDNA library (Marks et al., 2009) and the characterization of olivetol synthase (OLS), which shares 99% protein identity to our PKS candidate (Taguchi et al., 2008; Taura et al., 2009). Some of the work conducted on the PKS candidate presented in this thesis occurred prior to any of these reports on type III PKSs were published, and at the time this work was started, the PKS candidate had already been successfully subcloned, heterologously expressed, and purified by Ryan Taschuk, a member of the Page laboratory.

Table 3.1 Transcripts detected in a *C. sativa* glandular trichome cDNA library

Descriptions are from BLASTX comparison against the Uniprot database. Candidate cyclases assayed in this work are highlighted in yellow and include CHI (Q6V7U9), DABB/OAC (Probable protein POP3) (Q9LUV2), and MLP (B5KVN9). The most abundant transcript was TKS (B1Q2B6), initially described as OLS/polyketide synthase, is highlighted in pink. THCA synthase, another protein involved in cannabinoid biosynthesis, is also highlighted in pink. Adapted from an earlier publication that used the same cDNA library as in this study (Stout et al., 2012).

Description	UniProt #	Organism	Contig size	% ESTs
Olivetol synthase / polyketide synthase	B1Q2B6	Cannabis sativa	339	3.70
BURP domain-containing protein	B2ZPK5	Solanum tuberosum	192	2.10
4-hydroxy-3-methylbut-2-enyl diphosphate reductase	B9RZD3	Ricinus communis	185	2.02
Malic enzyme	B9RQE8	Ricinus communis	182	1.99
PAR-1c protein	Q43589	Nicotiana tabacum	87	0.95
Limonene synthase	A7IZZ1	Cannabis sativa	65	0.71
Delta12-oleic acid desaturase	Q6RXX0	Euphorbia lagascae	62	0.68
D12 oleate desaturase	Q41305	Solanum commersonii	53	0.58
Putative chalcone isomerase	Q6V7U9	Solanum lycopersicum	51	0.56
Non-specific lipid-transfer protein	Q850K8	Vitis hybrid	49	0.54
Tetrahydrocannabinolic acid synthase	Q8GTB6	Cannabis sativa	48	0.52
Isopentenyl pyrophosphate isomerase	Q6EJD1	Pueraria lobata	48	0.52
Elongation factor 1-alpha	B9SPV9	Ricinus communis	46	0.50
Cystathionine gamma-synthase	B9RYU1	Ricinus communis	44	0.48
Probable protein Pop3	Q9LUV2	Arabidopsis thaliana	41	0.45
Cytochrome P450	B9SG22	Ricinus communis	39	0.43
Alpha-pinene synthase	A7IZZ2	Cannabis sativa	38	0.41
Monoterpene synthase	B6SCF4	Humulus lupulus	35	0.38
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	A9ZN14	Hevea brasiliensis	33	0.36
ATP-citrate synthase	B9SHC9	Ricinus communis	31	0.34
Sesquiterpene synthase	B6SCF5	Humulus lupulus	31	0.34
Acetoacetyl-CoA thiolase	Q2L8A7	Picrorhiza kurrooa	28	0.31
Fructose-bisphosphate aldolase	Q9M4M9	Persea americana	28	0.31
Translationally-controlled tumor protein	B9RT61	Ricinus communis	28	0.31
Sesquiterpene synthase	B6SCF5	Humulus lupulus	27	0.29
Isoflavone reductase related protein	O81355	Pyrus communis	26	0.28
Short-chain dehydrogenase	B9RXD8	Ricinus communis	26	0.28
Salutaridinol 7-O-acetyltransferase	B9RML2	Ricinus communis	26	0.28
Lipoxygenase	C4NZX3	Camellia sinensis	25	0.27
Cysteine protease	A9UFX8	Vitis vinifera	25	0.27
Dessication-related protein	Q8LAU8	Arabidopsis thaliana	25	0.27
Anthranilate N-benzoyltransferase protein	B9S035	Ricinus communis	25	0.27
Hevein-like protein	Q944B8	Sambucus nigra	23	0.25
Pathogenesis related protein PR10	B5KVN9	Prunus domestica	23	0.25
Delta12-oleic acid desaturase	Q6RXX0	Euphorbia lagascae	22	0.24
Acetyl-CoA synthetase	B9RE68	Ricinus communis	22	0.24
Pyruvate decarboxylase	B9SWY1	Ricinus communis	21	0.23

In this chapter, experiments were designed to characterize the type III PKS, which was named tetraketide synthase (TKS), and set out to determine if an unknown chemical or biological factor played an accessory role in the biosynthesis of OA. Another goal of these experiemtns was to develop analytical techniques capable of detecting products from TKS assays performed *in vitro.* Recombinant TKS was assayed in a variety of buffers, pH ranges, and substrate concentrations to determine the optimal conditions for tetraketide formation. Rabbit anti-TKS polyclonal antibodies were also generated from purified TKS and used in co-immunoprecipitation and protein pull-down screens in an attempt to discover accessory proteins putatively involved in OA cyclization. These screening methods required a protein solution derived from isolated trichomes, which was assayed prior to any co-immunoprecipitation attempts. Recombinant TKS was also assayed in the presence of candidate proteins that were potentially involved in the cyclization of linear tetraketides. This chapter focuses on TKS and trichome assay results that collectively led to the discovery of OAC.

3.2 Materials and Methods

3.2.1 Tetraketide Synthase Assay Methods

All compounds used throughout this work were purchased (Sigma), except for PDAL, HTAL, and OA which were synthesized by Ms. Nikki Theaker using previously published methods (Horper and Marner, 1996).

3.2.2 The Expression and Purification of TKS

An *E. coli* BL21 (DE3) transformant 472-28A harboring pRT31A6 (*i.e.* TKS in pHIS8GW) was provided by Dr. Jonathan Page. The culture was streaked onto an LB plate, grown overnight at 37°C, and a single colony used to inoculate 5 mL of LB. Following overnight culture at 37°C, a

1 mL aliquot was transferred to start a larger overnight 50 mL culture. Frozen stocks were prepared by adding 100 µL of culture to 500 µL of 80% glycerol (v/v) and storing them at -80°C. The 50 mL culture was added to 1 L of AUTOINDUCTION MEDIUM (1.2% tryptone, 2.4% yeast extract, 0.0125% glucose, 0.05% α-lactose monohydrate, 50 mM glycerol, 50 mM Na_2HPO_4. 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 1 mM $MgSO_4$), and grown overnight at 30°C. The overnight culture was further induced by adding IPTG to a final concentration of 0.5 mM, and was grown for an additional 4 h at 30°C. The culture was pelleted at 10,000 g for 20 min at 4°C, and then transferred to a 50 mL tube before it was stored overnight at -80°C. All the media described contained 50 µg/mL kanamycin as a selective agent. Cultures were shaken at 200 rpm. Frozen pellets were thawed on ice for 30 min in the presence of 50 mL LYSIS BUFFER (50 mM TRIS-HCl [pH 7], 500 mM NaCl, 2.5 mM imidazole, 10% glycerol (v/v), 10 mM β-mercaptoethanol, 750 µg/mL lysozyme (Sigma)). The solution was homogenized using a vortexer, then sonicated on ice for 2 x 2 min bursts using a LabsonicU sonicator (B. Braun Biotech International). The cell suspension was centrifuged for 20 min at 12,000 g at 4°C and the supernatant was transferred to a 50 mL tube. In preparation of immobilized metal affinity chromatography (IMAC), a 1 mL suspension of Talon resin (Clontech) equilibrated in WASH BUFFER (50 mM TRIS-HCl [pH 7], 500 mM NaCl, 10 mM imidazole, 10% glycerol (v/v), 10 mM β-mercaptoethanol) was added to the cell suspension and then tumbled at 4°C for 40 min. Talon resin was pelleted by centrifugation for 1 min at 1,000 g, and the supernatant was discarded. Talon resin was washed five times using 10 mL of wash buffer each wash, where each wash was removed by centrifugation for 1 min at 1,000 g. The washed Talon was transferred to a 10 mL gravity flow column and rinsed with 10 mL wash buffer. Purified TKS was eluted using

10 mL ELUTION BUFFER (50 mM TRIS-HCl [pH 7], 150 mM NaCl, 100 mM imidazole, 10% glycerol (v/v), 10 mM β-mercaptoethanol), and concentrated to 1 mL using a YM10 centrifugal filter device (Millipore). The protein solution was desalted and the buffer exchanged by suspending it in 5 mL of STORAGE BUFFER (20 mM HEPES [pH 7.5] containing 25 mM NaCl, 10% glycerol (v/v), and 5 mM dithiothreitol (DTT)) and centrifuging the filtration device until the solution was concentrated to 1 mL. This dilution process was repeated two more times. Purification work was conducted at 4°C. Enzyme preparations were stored at -20°C.

3.2.3 TKS Assays

Unless stated otherwise, 50 µL reactions were conducted using 25 µg of TKS, 0.2 mM hexanoyl-CoA, 0.6 mM malonyl-CoA, and 5 mM DTT. Reactions were normally extracted twice using 250 µL ethyl acetate. Extracts were dried under vacuum and dissolved in 30 µL RESUSPENSION SOLUTION consisting of (v/v) 70% water, 30% acetonitrile, and 0.05% formic acid. Product analysis was generally conducted by liquid chromatography coupled to mass spectrometry (LCMS) and was normally performed using 10 µL injections of sample in a Waters Alliance 2695 separation module (Waters) attached to a 2996 photodiode array (PDA) detector (Waters), a 3100 single quad-mass detector (Waters), and a Symmetry C18 3.5 µm (2.1 x 100 mm) column (Waters). Products were monitored over 13 min isocratic runs using LCMS RUNNING SOLVENT consisting of (v/v) 63% water, 37% acetonitrile, and 0.05% formic acid. Flow rate was maintained at 0.2 mL/min with a 30°C column temperature. Results were processed using the Empower2 software suite (Waters). Assay products (HTAL, PDAL, OA, and olivetol) were detected by both LC-PDA and MS methods. Products detected by LC-PDA at 270 nm. MS was set to ES^+ mode where three selective ion monitoring (SIR) channels were used to

detect ions of m/z 181 (cone voltage = 25 V, dwell = 0.010 s), m/z 183 (cone voltage = 40 V, dwell = 0.010s), and m/z 225 (cone voltage = 25 V, dwell = 0.010 s). Ions were selected for the specific detection of olivetol (180 g/mol), PDAL (182 g/mol), and both HTAL and olivetolic acid (224 g/mol).

Some analyses were conducted solely by high performance liquid chromatography (HPLC) that used a Waters Alliance 2695 separation module (Waters) and 2996 photo array detector (PDA) (Waters) and a SunFire C18 10 µm (4.6 mm x 150 mm) (Waters). During HPLC analysis, samples were resuspended in 30 µL methanol, where 10 µL injections were monitored over 15 min gradient runs starting with (v/v) 59.9% water, 40% acetonitrile, and 0.1% trifluoroacetic acid and ending with 29.9% water, 60% acetonitrile and 0.1% trifluoroacetic acid. Flow rate was maintained at 1 mL/min with a 30°C column temperature. Reaction products were monitored at 270 nm, and results were processed using the Empower2 software suite. TKS reaction products were quantified using standard curves generated from 100 µM stock solutions prepared in LCMS running buffer. Solutions were diluted and 5 µL injections were used to construct 25 to 500 pmol standard curves following LCMS analysis (Appendix A). All curves presented a R^2 value ≥ 0.998 when analyzed by linear regression.

3.2.3.1 In the Presence of ^{14}C Malonyl-CoA and Hexanoyl-CoA

Aliquots of 25 µg TKS were assayed in 50 µL solutions containing 70 mM HEPES [pH 7.0], 20 µM malonyl-CoA, 20 µM hexanoyl-CoA and 1 µL of malonyl(2-^{14}C)-CoA (American Radiolabeled Chemicals Inc., 0.1 mCi/ml, specific activity 55 mCi/mmol) for 30 min at 30°C. Reactions were stopped and extracted using 200 µL ethyl acetate. Extracts were dried under vacuum, then resuspended in 10 µL of ethyl acetate before being spotted on a HPTLC-RP18F

plate (Analtech). The plate was developed over 60 min using a solvent consisting of 75% methanol, 24% water, and 1% acetic acid. The plate was analyzed by radio-TLC scanning using a LS6500 scintillation counter (Beckman Coulter) and by phosphorimaging using a Storm 840 imaging station (Amersham Biosciences).

3.2.3.2 In the Presence of Malonyl-CoA and Hexanoyl-CoA

Reactions were conducted using 20 mM HEPES [pH 7.0] at 20°C for 60 min. Extracts were dried under vacuum and then resuspended in 20 µL of resuspension solution. Assays were performed in triplicate.

3.2.3.3 In the Presence of Different Buffers and ph

Reactions used 20 µg of TKS and were conducted for 30 min at 25°C in 20 mM buffer, 0.2 mM malonyl-CoA, 0.2 mM hexanoyl-CoA, and 5 mM DTT. Buffers included 20 mM MES [pH 5.5, 6.0, 6.5], phosphate [pH 6.0, 6.5, 7.0, 7.5, 8.0], HEPES [pH 7.0, 7.5, 8.0], and TRIS [pH 7.0, 7.5, 8.0, 8.5, 9.0]. Assays were performed in triplicate and analyzed by HPLC.

3.2.3.4 In the Presence of N-(2-acetamido)iminodiacetic acid (ADA) Buffer

Reactions used 100 pmol (~5 µg) of TKS and were conducted in 20 mM ADA buffers [pH 5.7, 6.0, 6.2, 6.4, 6.6, 6.8, 7.0, 7.2, 7.4, 7.6, 7.8]. Due to the large number of samples, reaction times were extended to 90 min at 20°C to ensure that there was time enough to start/stop samples. Extracts were resuspended in 20 µL resuspension solution and used 5 µL injections during LCMS analysis. Assays were performed in triplicate.

3.2.3.5 With Variable Temperatures

TKS temperature assays used 20 µg TKS and were conducted in 50 mM HEPES [pH 7.5], 0.2 mM malonyl-CoA, and 0.2 mM hexanoyl-CoA. Reactions were incubated for 30 min at 10, 20, 30, 40, or 50°C. Reactions were performed in triplicate and were analyzed by HPLC.

3.2.3.6 With Variable Substrate Concentrations

Assays were performed in 100 mM HEPES [pH 7.5] containing 5 mM DTT, and either 0.2 mM hexanoyl-CoA and 0.1, 0.2, 0.5 or 1.0 mM malonyl-CoA, or 0.2 mM malonyl-CoA and 0.1, 0.2, 0.5, or 1.0 mM hexanoyl-CoA. Reactions were performed at 40°C for 30 min in triplicate and were analyzed by HPLC.

3.2.3.7 In the Presence of DTT

Two sources of TKS were used in assays, both of which were derived from the same culture. Following the expression of TKS, 472-28A strains were divided into two portions so to purify proteins in either the presence or absence of β-mercaptoethanol (βME) and DTT in purification buffers. TKS purified in the presence or absence of reducing agents was used to assay the effect of DTT in assays. Reactions contained 100 mM HEPES [pH 7.5], 0.2 mM malonyl-CoA, 0.2 mM hexanoyl-CoA, and either 0, 5, 10, 20, or 100 mM DTT. Reactions proceeded for 90 min at 40°C before being extracted twice with 200 µL of ethyl acetate. Reactions were performed in triplicate and were analyzed by HPLC.

3.2.4 Cannabis Trichome Preparation and Trichome Extract Assays

Two different trichome extracts were used during assays; one derived from non drug-type cannabis (USO-31 hemp), and the other from drug-type cannabis. TRICHOME EXTRACTION

BUFFER consisted of 100 mM HEPES [pH 7.5], 3 mM EDTA, 500 mM sucrose, 10 mM DTT, and 1,000x diluted protease inhibitor mixture set III (Calbiochem).

3.2.4.1 Trichome Isolation

Female cannabis plants provided by Dr. Page had been grown in a controlled environment chamber at 18 h day lengths. Flowering bracts (60 g) were removed from the plants and dissected before soaking them for 1 h in ice water. Trichome separation was achieved using an adapted Beadbeater method (Gershenzon et al., 1992). The flowering material was moved to a 950 mL vessel (BeadBeater, BioSpec Products) containing 120 mL of 0.5 mm glass beads, 35 g XAD-4 and 700 mL of TRICHOME ISOLATION BUFFER (25 mM HEPES pH 7.0, 200 mM sorbitol, 10 mM sucrose, 0.6% (w/v) methylcellulose and 1% (w/v) polyvinylpyrrolidone (40,000 g/mol) (PVPP)). The mixture was agitated using a teflon rotor for 1 min before filtering it through a 350 µ mesh. The tissue was then washed with 250 mL of trichome isolation buffer which did not contain methyl cellulose or PVPP. All filtrates were combined and run through a 105 m mesh, and then again through a 20 m mesh filter prior to trichome collection. The trichome cells were rinsed using trichome isolation buffer and dried using a Kimwipe (Kimberly-Clark). Trichomes were stored at -80°C. Yield was *ca.* 0.9 g from 60 g of flowers.

3.2.4.2 Cannabis Trichome Extracts

A 100 mg sample of isolated drug-type trichomes was homogenized in the presence of 10 mg PVPP and 200 µL trichome extraction buffer on ice, followed by a 20 min centrifugation at 15,000 g at 4°C. The supernatant was removed, and trichomes were extracted again using another 200 µL of extraction buffer. Protein extracts were combined and desalted using a 2 mL Zeba desalting spin column (Thermo Scientific) equilibrated with 100 mM HEPES [pH 7.5].

Protein concentration of the extracts was determined using an RC/DC protein assay as directed by the manufacturer (Bio-Rad). Protein extracts were assayed using 25 µg (ca. 45 µL) of protein for 30 min at 40°C in a final volume of 55 µL containing 0.4 mM hexanoyl-CoA and 0.4 mM malonyl-CoA or simply 0.4 mM hexanoyl-CoA and water (no malonyl-CoA negative control). Samples were prepared for HPLC analysis as described in Section 3.1.3. Injections of 20 µL were analyzed by HPLC.

Scaled-up USO-31 hemp trichome assays used 195 µg of protein (ca. 350 µL) from protein extracts prepared as described above. Extracts were incubated in the presence of 0.25 mM malonyl-CoA and 0.25 mM hexanoyl-CoA for 2 h at 40°C. The 400 µL solution was extracted twice using 2 x 1 mL ethyl acetate. A single 20 µL injection was performed during HPLC analysis.

3.2.4.3 The Effect of Recombinant TKS on Trichome Extract Activity

Trichome extracts were prepared as described in Section 3.1.4.2. A total of five reactions were conducted including 1) recombinant TKS (rTKS), 2) trichome extract, 3) trichome extract in the presence of rTKS 4) trichome extract lacking hexanoyl-CoA (negative control), and 5) rTKS lacking hexanoyl-CoA (negative control). The 55 µL TKS reaction was conducted by incubating 50 µg of rTKS in the presence of 80 mM HEPES [pH 7.5], 0.4 mM malonyl-CoA, and 0.4 mM hexanoyl-CoA. The trichome extracts were assayed by incubating 200 µg of protein from trichome extract in the presence of 0.5 mM malonyl-CoA and 0.25 mM hexanoyl-CoA. A negative control for this reaction replaced hexanoyl-CoA with water. The TKS/trichome assay included 50 µg of TKS with 200 µg protein from trichome extracts. Reactions were incubated for 2 h at 40°C before being extracted using five volumes of ethyl acetate. Samples were prepared

for HPLC analysis as previously described (Section 3.1.3) and used 20 µL injections during analysis.

3.2.4.4 Rabbit Antibodies Directed Against TKS

Antibodies were prepared by the National Research Council Canada, Biotechnology Research Institute (NRC-BRI) using their standard immunization protocols for the development of polyclonal antibodies, which required the inoculation of two rabbits each with 350 µg of purified recombinant TKS. Recombinant TKS was prepared and purified as outlined in Section 3.1.1. See Appendix B for details on rabbit injections and bleed times.

3.2.4.5 Ammonium Sulfate Precipitation of Trichome Proteins

Trichome extracts were prepared as previously described in Section 3.1.4.2. A saturated ammonium sulfate solution was prepared by stirring 7.5 g of ammonium sulfate in 10 mL of water. A 1 mL aliquot of extract was tumbled in the presence of 20% ammonium sulfate for 1 h at 4°C, followed by centrifugation at 15,000 g for 20 min. The supernatant was removed and the 20% ammonium sulfate pellet was stored at -20°C. The supernatant was then tumbled in 40% ammonium sulfate for 1 h at 4°C, centrifuged at 15,000 g for 20 min, the supernatant removed, and the pellet stored at -20°C. In a similar fashion, the remaining supernatant was precipitated using 60% and 80% ammonium sulfate. The remaining supernatant (*i.e.* post-80% ammonium sulfate treatment) was desalted using a 5 mL PD10 column (GE Healthcare Biosciences) and stored at -20°C. Pellets were later thawed on ice and resuspended in 100 µL of 100 mM HEPES [pH 7.5]. Since there was some difficulty in reconstituting the 20% pellet in this volume, an additional 100 µL of buffer was used (total 200 µL). Solutions were concentrated using a YM10 concentrator to ca. 150 µL. An aliquot of 3 µL from each solution was added to 2 µL SAMPLE

LOADING BUFFER (50 mM TRIS [6.8], 2% SDS (w/v), 10% glycerol (v/v), 1% βME (v/v), 0.02% bromophenol blue (w/v)) and incubated for 5 min at 95°C before being analyzed by SDS-PAGE and Western blotting using 10% gels run at 200 V for 45 min. One gel was stained in Bio-Safe Coomassie stain for 1 h, and destained overnight in dH_2O. The other gel was electrotransferred to a PVDF membrane using 15 V for 30 min. The transferred PVDF membrane was blocked with a 10% BSA/phosphate buffered saline (PBS) (Sigma-Aldrich) solution for 1 h and washed twice in PBS for 10 min, then 2 more times each time for 10 min in PBS containing 0.1% Tween 20. A 1:1,000 primary antibody solution was prepared by adding 20 µL of rabbit antibodies (see Section 3.1.4.5) to 20 mL of 1% BSA in PBS, which was used to treat the PVDF membrane on an orbital shaker for 1 h. A 1:20,000 secondary antibody (goat anti-rabbit conjugated with alkaline phosphatase) was prepared in 1% BSA/PBS and allowed to react with the membrane for 40 min on an orbital shaker. The membrane was washed as above and investigated by chemiluminescence using 1 mL of substrate from the Immun-Star goat anti-rabbit IgG (H + L) Detection Kit (Bio-Rad). A 30 s exposure on Kodak Biomax XAR film (Sigma-Aldrich) was used to detect anti-TKS antibodies in protein extract solutions.

3.2.4.6 Activity of Ammonium Sulfate Trichome Protein Fractions

Trichome protein extracts were assayed by incubating 50 µL of the resuspended solutions (0%, 20%, 40%, 60%, 80%, and remaining supernatant) with 0.3 mM malonyl-CoA and 0.3 mM hexanoyl-CoA. A TKS positive control was used by assaying 25 µg of TKS in the presence of 0.2 mM malonyl-CoA and 0.2 mM hexanoyl-CoA. A 50 µL aliquot from the 20% fraction was assayed in the absence of hexanoyl-CoA and used as a negative control. Reactions were started

by adding the appropriate fraction (or TKS), and proceeded for 60 min at 40°C. Reaction mixtures were prepared for HPLC as described in Section 3.1.3.

3.2.4.7 Co-Immunoprecipitation of TKS

Experiments were conducted with a Co-immunoprecipitation Kit (ThermoScientific) using 4 mL of rabbit anti-TKS antibodies (see Section 3.1.4.5) purified as directed by the manufacturer. Purified IgG was desalted using two 5 mL D-Salt Excellulose Desalting Columns equilibrated with 100 mM PBS [pH 7.3]. Purified antibodies were bound to a Cyanogen Bromide-Activated Sepharose 4B (CNBR) column. Purified IgG (50 µg) was used during column preparation. Two other columns were also prepared but did not include the use of antibodies, and were used as 1) a negative control, 2) a pre-binding column to minimize binding that could result from non-specific interactions between the resin and the sample. Resin and antibodies were incubated for 90 min at RT on an orbital shaker. The resin columns were washed six times, and rinsed once with 1x coupling buffer. Columns were washed three more times using 200 µL of wash buffer before further use. Initially, the columns were tested for their efficiency in binding recombinant TKS. Two samples of crudely purified TKS (100 µg) were used to determine the specificity of the prepared columns. The 0.5 µL samples were diluted to 400x (200 µL) before applying them to either column 1) containing control resin, or 2) pre-bind column. Following centrifugation, the flow-through was either 1) saved for further analysis in the case of the negative control, or 2) passed through the column containing IgG-bound resin which was incubated 2 h on a shaker at RT. All columns were washed four times with wash buffer, followed by three elutions using 20 µL of elution buffer. All fractions were saved for analysis by SDS-PAGE. 20 µL from each fraction were incubated for 3 min at 98°C in the presence of 5 µL of the provided SDS-PAGE

loading buffer. Samples from each column were loaded onto separate 10% acrylamide gels and subjected to electrophoresis for 30 min at 200 V. Gels were stained using BioSafe coomassie blue (Bio-Rad) for 1 h, and washed overnight using dH_20.

Trichome extracts were prepared as previously described in Section 3.1.4.2. A total of 500 µL of trichome extract was divided into two equal portions and applied to either 1) negative control IP column, or 2) pre-bind IP column. The pre-bind sample was incubated for 60 min at 4°C upright on a shaker, and then centrifuged at 1,500 g for 90 s. The flow-through was collected. The pre-bind column was washed four times, then eluted using 3 x 1 mL elution buffer. Flow-through from the pre-bind column was applied to the antibody-bound resin IP column and allowed to incubate overnight at 4°C alongside the negative control upright on a shaker. Columns were later washed four times using 200 µL of IP wash/lysis buffer, followed by centrifugation of 1,500 g for 90 s after each wash. A 30 µL volume of elution buffer was added to each column, and after 5 min columns were centrifuged at 1,500 g for 5 min. Two more elutions using 20 µL elution buffer were performed, before columns were washed and placed in 1x coupling buffer for storage at 4°C. A total of 20 µL from each column fraction (flow-through, washes, and eluted fractions) were heated with 5 µL of sample loading buffer for 5 min at 98°C. Samples were analyzed by SDS-PAGE as described in Section 3.1.4.5.

3.2.4.8 Protein Pull-Downs of Trichome Protein Extracts

This experiment used bacterial lysate that had already undergone TKS purification as previously described (Section 3.1.1.). A second round of TKS purification was performed using the lysate, which was tumbled with 1 mL Talon beads at 4°C for 40 min. The beads were washed five times using 10 mL wash buffer and transferred to a 10 mL gravity flow column. A 200 µL aliquot of

trichome extract that had been previously prepared (Section 3.1.4.2) was thawed, transferred to the 10 mL gravity flow column and shaken for 1 h at RT. The gravity flow column was washed three times using 10 mL wash buffer each time. Proteins were eluted using 10 mL elution buffer. The solutions were concentrated to a 200 µL volume using a YM10 centrifugal filter device (Millipore), and desalted by resuspending the solution in 3 mL of 100 mM HEPES [pH 7.5], and then concentrating to a 200 µL volume. Eluted proteins were compared by SDS-PAGE to a sample of purified TKS obtained from the first round of purification, and to untreated trichome extracts. Samples included (separate) 2 µg TKS, 3 µg of eluted proteins, 6 µg of eluted proteins, 1 µg of trichome protein, and 2 µg of trichome protein, which were each heated to 95°C for 5 min in the presence of sample loading buffer prior to loading them onto a 10% acrylamide gel. Proteins were analyzed by SDS-PAGE as previously described in Section 3.1.4.5.

3.2.5 Cloning and Assaying of Candidate Cyclases

3.2.5.1 Cloning and Expression of Candidate Cyclases

Chi-like, *mlp-like*, and *oac* (Table 3.1) were PCR-amplified using the primers described in Table 3.2 and conditions outlined in Table 3.3. Whereas *chi* and *mlp* were cloned using Phusion High Fidelity DNA Polymerase (Finnzymes), each *oac* construct was amplified using Pfu Ultra II DNA Polymerase (Stratagene). Amplicons were A-tailed by incubating with Taq polymerase (Invitrogen) in the presence of 5 mM dATP for 30 min at 72°C and purified using a QIAquick Spin PCR Purification Kit (Qiagen). Following electrophoresis in 0.8% agarose, single bands were excised and cleaned using a MinElute PCR Purification Kit (Qiagen) before being either ligated into the linear pCR8GW entry vector (Invitrogen) or, in the case of OAC, cloned directly into pET100 (see Table 3.3). Plasmids were used to transform *E. coli* One Shot Top10

Table 3.2 Primers used during cloning of CHI-like, MLP-like, and OAC constructs

Gene	Forward Primer	Reverse Primer
chi-like	ATGGCAACCTCAGATGGAAGC	TTAGAAGAGAGAAAGCATGGA
mlp-like	ATGGGTGTTTTCACTTACG	TTAGTTGTAAGCTTCAGGGTG
oac/DABB[1]	CACCATGGCAGTGAAGCATTTGATTGTATTG	CTTTCGTGGTGTGTAGTCAAAAATGAGAAG
oac/DABB[2]	ATGGCAGTGAAGCATTTGATT	CTACTTTCGTGGTGTGTAGTC

1 - Original pET100 based OAC construct.
2 - pHIS8GW based OAC construct.

Table 3.3 Details regarding the creation of CHI-like, MLP-like, and OAC constructs

Gene	Template	Melting[1]	Annealing[1]	Elongation[1]	Destination	Selection[2]
chi-like	CL8contig1[3]	98°C/30s	60°C/10s	78°C/15s	pHIS8GW	Kan^R
mlp-like	pJS16[3,4]	98°C/30s	55°C/30s	72°C/30s	pDEST17	Amp^R
oac/DABB[5]	CL2contig1[3]	95°C/20s	55°C/20s	72°C/60s	pET100	Amp^R
oac/DABB[6]	pET100-OAC	95°C/20s	50°C/20s	72°C/18s	pHIS8GW	Kan^R

1 - The reactions were cycled 30-35 times before ending with 7 min at 72°C.
2 - Antibiotic resistance markers include Kanamycin (Kan^R) and Ampicillin (Amp^R).
3 - Originally derived from cDNA
4 - An out-of-frame mlp-like construct originally cloned by Jake Stout from trichome cDNA template.
5 - Original pET100 based OAC construct.
6 - pHIS8GW based OAC construct

Competent Cells (Invitrogen), and transformants were plated onto LB containing 50 µg/mL spectinomycin (or 100 µg/mL ampicillin for pET100 construct), and grown overnight at 37°C. Colonies were selected for each construct and grown overnight under selective pressure at 37°C. Plasmids were isolated and sequenced as previously described (Section 3.1.1). Following sequence confirmation, genes were transferred from the entry vector into the appropriate destination vector by LR reaction following the manufacturer's directions (Invitrogen) (see Table 3.3), and the resulting plasmids transformed into *E. coli* Rosetta2 BL21(DE3)pLysS Competent

Cells (Novagen Biosciences). Transformants were cultured with the appropriate antibiotic, cultures were induced, and proteins purified as previously described (Section 3.1.1).

3.2.5.2 Assaying TKS in the Presence of Candidate Cyclases

Reactions used 100 pmol of TKS (ca. 5 µg) and either 500 pmol of CHI-like, MLP-like, OAC, or in the case of negative controls, no protein, and were conducted for 1 h at 30°C. Samples were prepared and analyzed by LCMS in triplicate as described in Section 3.1.3.

3.3 Results

3.3.1 Tetraketide Synthase (TKS)

TKS contains 385 amino acids (AAs) and is a 43 kDa protein with a 6.04 pI (47 kDa and 6.58 pI with histidine tag). A BLASTP homology search of the NCBI non-redundant database reveals TKS to be 100% identical to PKSG4 (annotated ACD76855) (Flores-Sanchez et al., 2010), and similar to several *Cannabis* homologs, including OLS (annotated BAG14339; 385 AAs; 99% identity to TKS in 385 AA overlap). These enzymes belong to a superfamily of condensing enzymes (cl09938) that contain a conserved CHS-like domain (cd00831) attributed to plant specific type III PKSs. TKS belongs to a subset of type III PKSs that do not produce chalcone and are markedly different than other cannabis CHSs previously reported in the literature (annotated AAL92879; 389 AA; 68.9% identity over a 379 AA subset with TKS) (Raharjo et al., 2004b), (annotated ACT10338; 392 AA; 66.0% identity in 377 AA overlap with TKS) (Marks et al., 2009). An alignment of known cannabis type III PKSs highlights differences in their amino acid sequence (Figure 3.1).

```
TKS          1 ---------MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRK
PKSG4        1 ---------MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRK
PKSG1        1 ---------MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRK
PKSG5        1 ---------MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRK
PKSG2        1 ---------MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRK
PKSF3        1 ---------MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRK
OLS          1 ---------MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEKFRK
CAN383       1 MASISVDQIRKAQRANGPATVLAIGTANPPTSFYQADYPDFYFRVTKNQHMTELKDKFKR
CHS2004      1 --MVTVEEFRKAQRAEGPATIMAIGTATPANCVLQSEYPDYYFRITNSEHKTELKEKFKR

TKS         52 ICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWG
PKSG4       52 ICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWG
PKSG1       52 ICDKSMIRKRNIFLNEEHLKQNPKLVEHDVQTLDARQDMLVVEVPKLGKDACAKAIKEWG
PKSG5       52 ICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWG
PKSG2       52 ICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWG
PKSF3       52 ICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWG
OLS         52 ICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWG
CAN383      61 ICEKTTIKKRHLYLTEDRLNQHPNLLEYMAPSLNTRQDMLVVEIPKLGKEAAMKAIKEWG
CHS2004     59 MCDKSMIRKRYMHLTEEILKENPNLCAYEAPSLDARQDMVVVEVPKLGKEAATKAIKEWG

TKS        112 QPKSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIA
PKSG4      112 QPKSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIA
PKSG1      112 QPKSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIA
PKSG5      112 QPKSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIA
PKSG2      112 QPKSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIA
PKSF3      112 QPKSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIA
OLS        112 QPKSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIA
CAN383     121 QPKSRITHLIFCSTNGVDMPGADYECAKLLGLSSSVKRVMLYQQGCHAGGSVLRIAKDLA
CHS2004    119 QPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLA


TKS        172 ENNKGARVLAVCCDIMACLFRGPSDSDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIF
PKSG4      172 ENNKGARVLAVCCDIMACLFRGPSDSDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIF
PKSG1      172 ENNKGARVLAVCCDIMACLFRGPSDSDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIF
PKSG5      172 ENNKGARVLAVCCDIMACLFRGPSDSDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIF
PKSG2      172 ENNKGARVLAVCCDMTACLFRGPSDSNLELLVGQAIFGDGAAAVIVGAEPDESVGERPIF
PKSF3      172 ENNKGARVLAVCCDIMACLFRGPSDSDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIF
OLS        172 ENNKGARVLAVCCDIMACLFRGPSESDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIF
CAN383     181 ENNKGARILTINSEITIGIFHSPDETYFDGMVGQALFGDGASATIVGADPDKEIGERPVF
CHS2004    179 ENNKGARVLVVCSEITAVTFRGPNDTHLDSLVGQALFGDGSAALIVGSDPIPEV-EKPIF

TKS        232 ELVSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNS
PKSG4      232 ELVSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNS
PKSG1      232 ELVSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNS
PKSG5      232 ELVSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNS
PKSG2      232 ELVSTGQTFLPNSEGTIGGHIREAGLMFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNS
PKSF3      232 ELVSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNS
OLS        232 ELVSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNS
CAN383     241 EMVSAAQEFIPNSDGAVDGHLTEAGLVYHIHKDVPGLISKNIEKSLVEALNPIGISDWNS
CHS2004    238 ELVSAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLNEAFKPLGISDWNS
```

```
TKS       292 IFWITHPGGKAILDKVEEKLDLKKEKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEG
PKSG4     292 IFWITHPGGKAILDKVEEKLDLKKEKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEG
PKSG1     292 IFWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEG
PKSG5     292 IFWITHPGGKAILDKVEEKLHLKKEKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEG
PKSG2     292 IFWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEG
PKSF3     292 IFWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEG
OLS       292 IFWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEG
CAN383    301 LFWIVHPGGPAILNAVEAKLHLKKEKMADTRHVLSEYGNMSSVSIFFIMDKLRKRSLEEG
CHS2004   298 LFWIAHPGGPAILDQVESKLALKTEKLRATRHVLSEYGNMSSACVLFILDEMRRKCVEDG

TKS       352 KSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY
PKSG4     352 KSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY
PKSG1     352 KSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY
PKSG5     352 KSTTGDGFEWGVLFGFGPGLTVETVVLRSVPINY
PKSG2     352 KSTTGDGFEWGVLFGFGPGLTVERVVLRSVPINY
PKSF3     352 KSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY
OLS       352 KSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY
CAN383    361 KSTTGDGFEWGVLFGFGPGLTVETIVLHSLAN--
CHS2004   358 LNTTGEGLEWGVLFGFGPGLTVETVVLHSVAI—
```

Figure 3.1 Alignment of TKS and other cannabis PKS homologs.

TKS was aligned with other reported cannabis PKSs using Clustal Omega (Sievers et al., 2011) and Boxshade software. The alignment members are PKSG1 (annotated ACD76853), PKSG2 (annotated ACD76854), PKSF3 (annotated ACD76852), PKSG4 (annotated ACD76855), PKSG5 (annotated ACD76856) (Flores-Sanchez et al., 2010), OLS (annotated BAG14339) (Taura et al., 2009), CAN383 (annotated ACT10338) (Marks et al., 2009), and CHS2004 (annotated ACT10338) (Raharjo et al., 2004b). Residues which differ from the consensus are depicted in blue if they share similarity with the consensus; residues that do not share similarity are depicted in red. TKS and PKSG4 share 100% identity.

3.3.2 Expression and Purification of TKS

TKS was purified to near homogeneity by exploiting a poly-histidine tag by IMAC using Talon. The expression and purification of TKS gave yields typically reaching 20 mg/L. Purified TKS was visualized by SDS-PAGE on a 10% acrylamide gel (Figure 3.2).

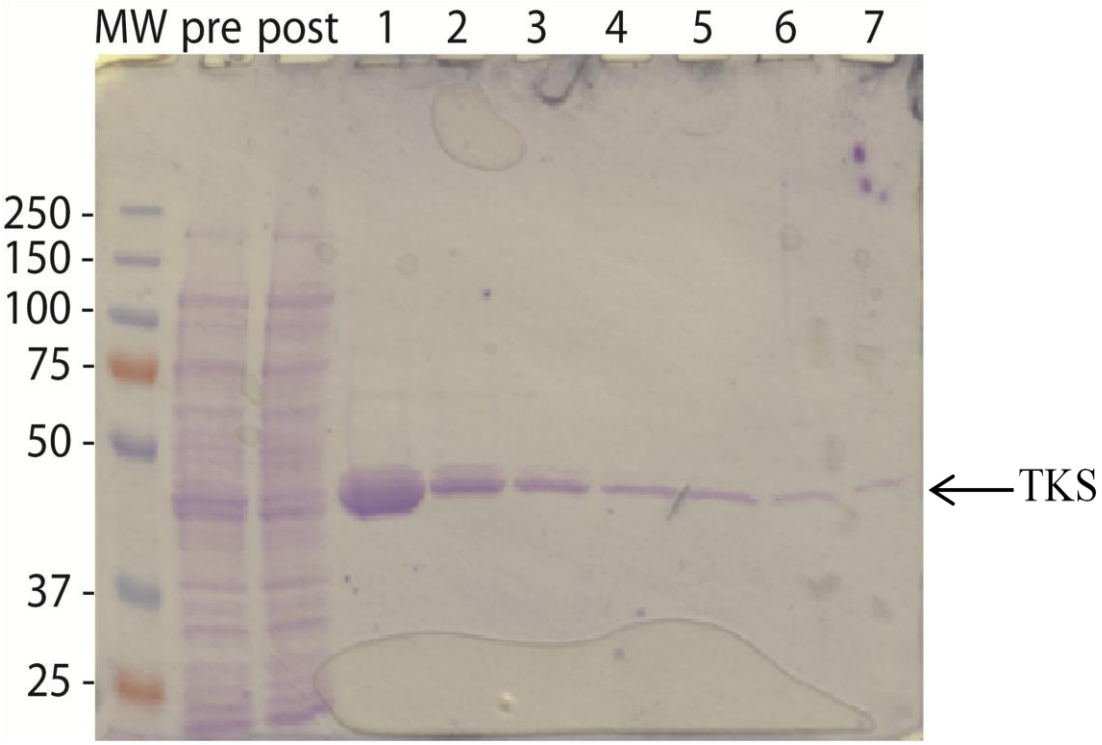


Figure 3.2 Expression and purification of TKS.

The SDS-PAGE analysis of a (MW) BioRad Dual Color protein marker and samples taken during the pre-induction (pre) and post-induction (post) stages of *E. coli* BL21 cell cultures harboring *CsTks*, and from 1 mL eluents of IMAC purified TKS from Talon resin (lanes 1 to 7). Purified TKS migrated at its estimated 46.5 kDa size.

3.3.3 TKS produces HTAL, PDAL, and olivetol

Different analytical methods were explored to determine which could best detect TKS assay products including the use of radio-labeled ^{14}C-malonyl-CoA and reverse-phase thin-layer chromatography (Figure 3.3). The analysis of assay products by reverse-phase thin-layer chromatography indicated the production of three compounds assumed to be HTAL, PDAL, and olivetol, although it was determined that the use of an HPLC, and later LCMS, proved a better option for analysis since it is capable of superior product resolution, quantification, and mass confirmation (Figure 3.4). The use of authentic standards (Sigma) allowed for the identification and quantification of the three compounds produced during TKS assays: hexanoyl triacetic acid lactone (HTAL), pentyl-diacetic acid lactone (PDAL), and olivetol, but no OA was produced (Figure 3.4). These results confirm that olivetol synthase activity as reported by Taura et al., (2009) and Marks et al., (2009) is also present in TKS. Standard and product LCMS traces, along with their UV spectra and masses are shown in Figure 3.4. The MS characteristics of HTAL, PDAL, OA, and olivetol were also analyzed by SIR is ES^+ mode (Section 3.1.3 for details) (Figure 3.5). Note that the results shown in Figures 3.4 and 3.5 are from assays conducted once LCMS methods were well developed, and earlier results (such as seen in Figures 3.13and 3.15), which were based only on HPLC analysis, have product retention times that slightly differ from later LCMS analysis. TKS assays accumulated PDAL, olivetol and HTAL in a ratio of 10:5:1, respectively. Although the ratio of triketide:tetraketide production varied during the intial characterization of TKS, the production of PDAL was always greater through all experiments performed.

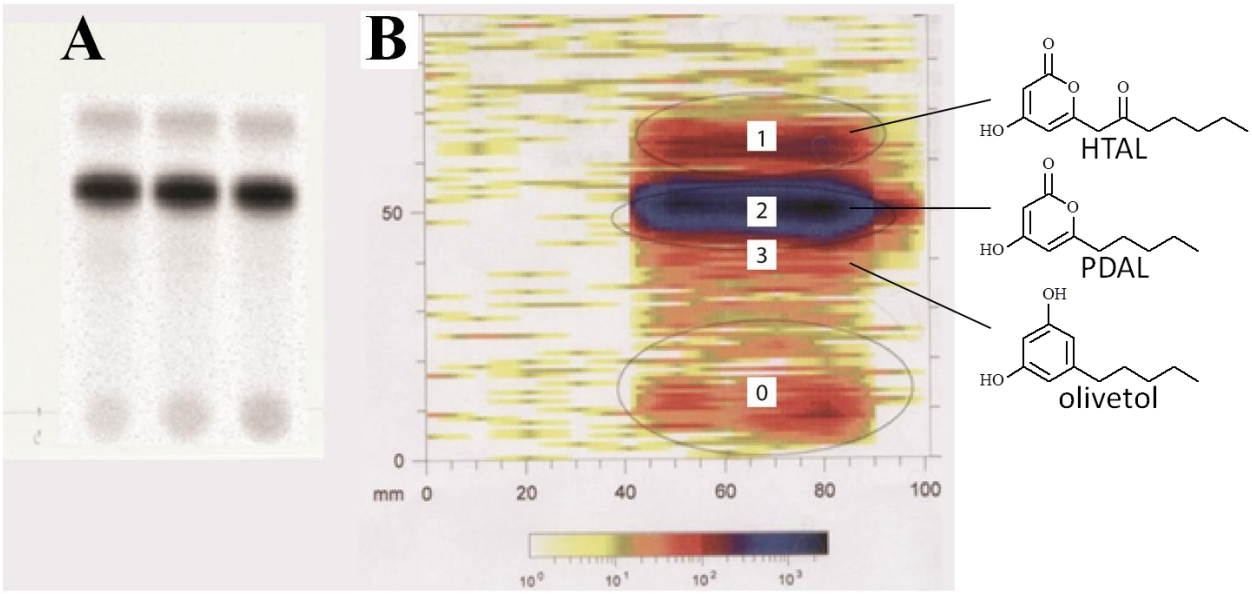


Figure 3.3 Reverse phase radio TLC analysis of TKS assay products.

Reverse-phase thin layer chromatography (TLC) was used to separate TKS assay products (HTAL, PDAL, and olivetol) which were analyzed by (A) phosphorimaging and (B) radioactive scanning. Products detected are assumed to be (1) HTAL, (2) PDAL, and (3) olivetol, where (0) is the origin. Further developments of these methods were abandoned for HPLC and LCMS methods which led to a better resolution of the products.

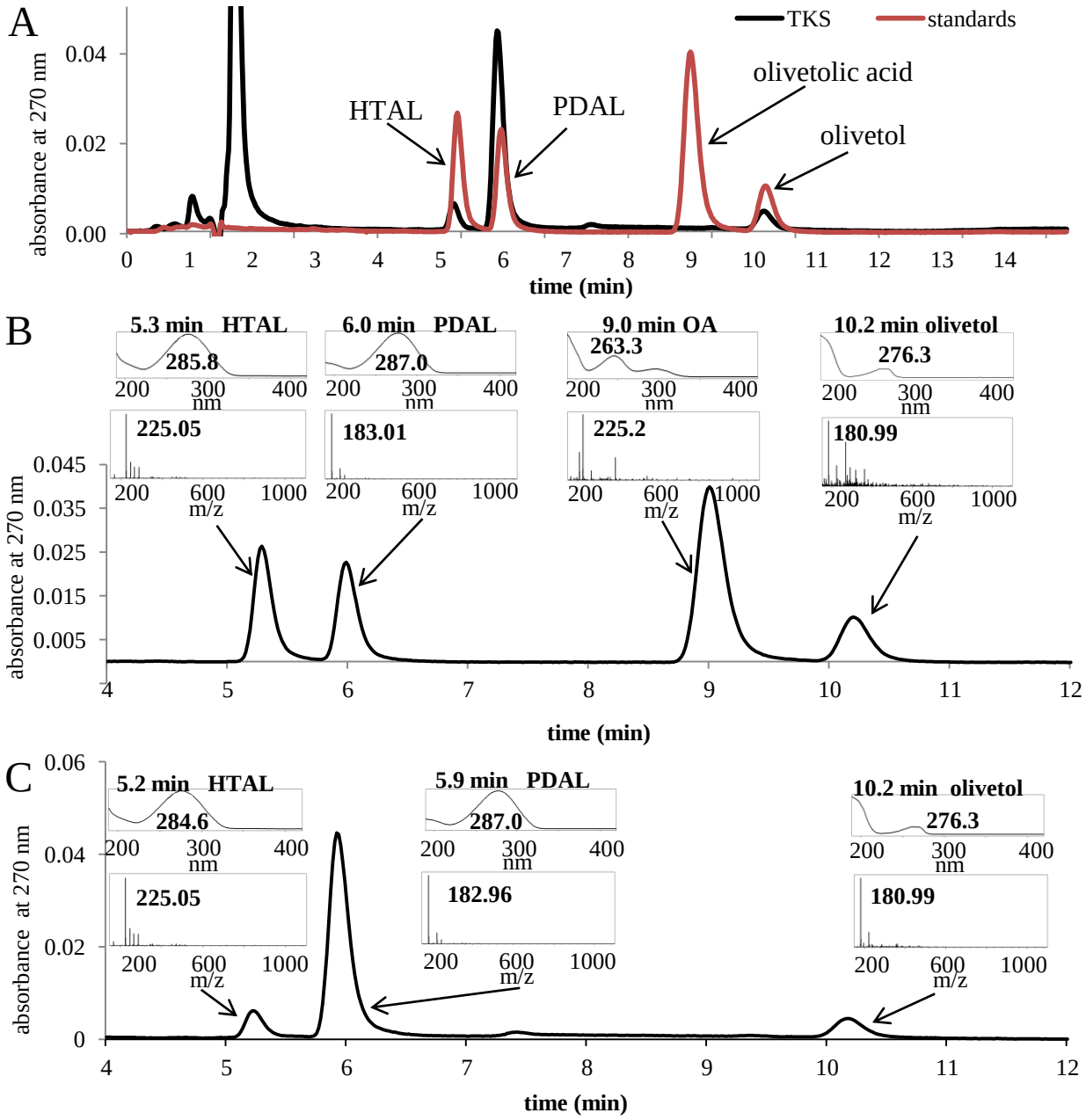


Figure 3.4 LCMS analysis of standards and TKS products.

(A) LC chromatograms for TKS assay products (black) and standards (red) are overlayed to emphasize the similarity in retention times, except for OA which was absent in TKS assays. (B) An LCMS chromatogram of the standards used in this study indicate the retention times, UV spectra and extracted ion profiles for HTAL (5.2 min; UV_{max} 285 nm; m/z 225), PDAL (5.6 min; UV_{max} 287 nm; m/z 183), OA (9.0 min; UV_{max} 263 nm; m/z 225.), and olivetol (10.2 min; UV_{max} 276 nm; m/z 181) and used to confirm the identities of (C) TKS assay products shown in a similar LCMS trace.

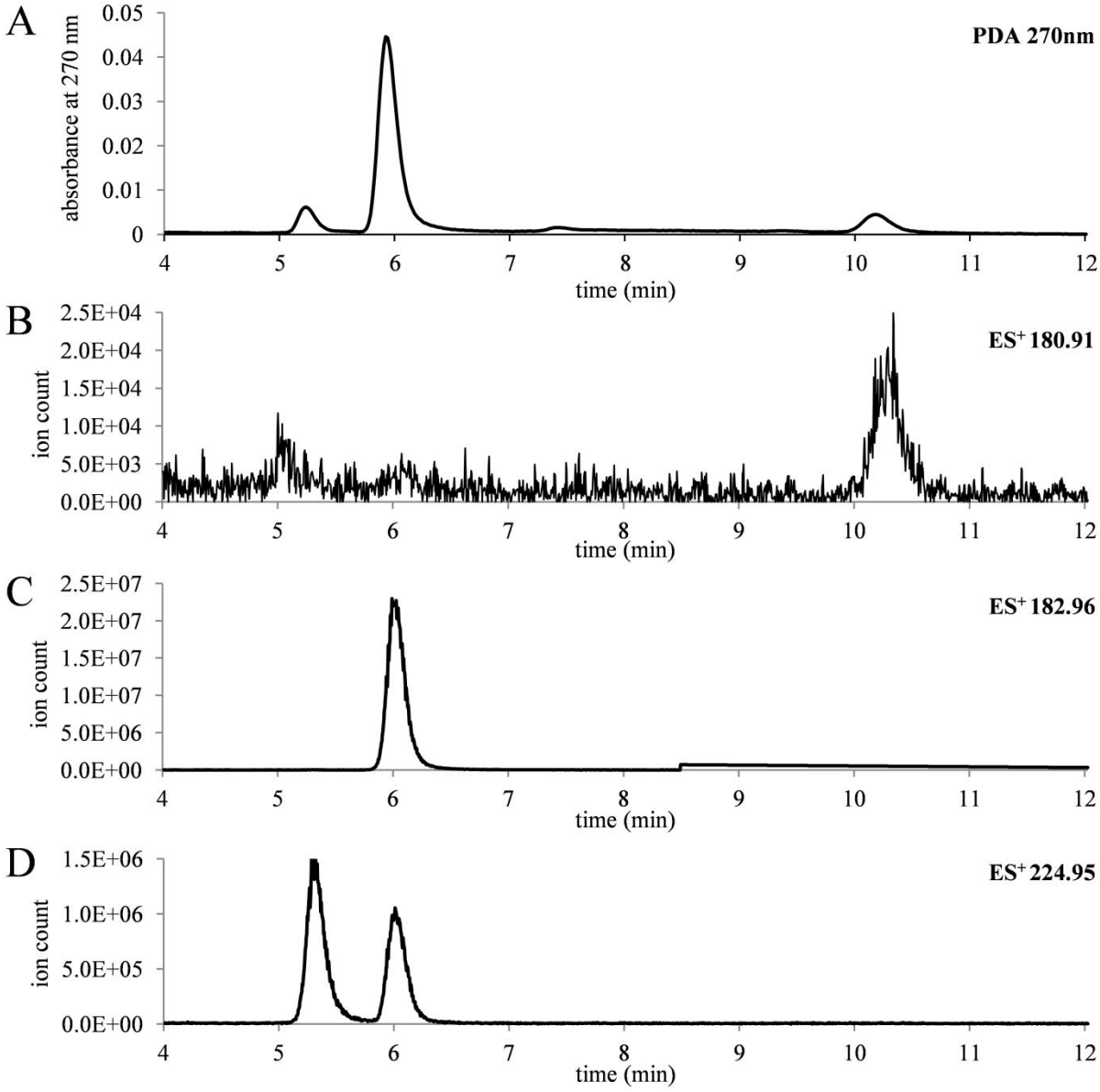


Figure 3.5 Selected ion monitoring of TKS assay products.

LCMS was used to confirm the identity of HTAL, PDAL and olivetol produced during TKS assays. (A) PDA results align well with selected ion monitoring in ES^+ mode which was capable of detecting (B) olivetol (10.2 min; m/z 180.91), (C) PDAL (5.9 min; m/z 182.96), and (D) HTAL (5.2 min; m/z 224.95). Another peak (5.9 min) detected during (D) selected ion monitoring at 224.95 m/z is attributed to acetonitrile:PDAL adduct formation.

3.3.3.1 Quantification of TKS products

Quantification of TKS reactions products (HTAL, PDAL, OA, and olivetol) was performed using the standard curves presented in Appendix A. These standard curves were used to determine product yields (pmol) from assays conducted throughout this study.

3.3.3.2 Buffers and pH

The effect of buffering agents and pH on TKS activity was examined by assaying TKS in a series of buffers that overlapped in pH ranges (Figure 3.6). Assay results indicate that although olivetol production increases at a lower pH (ca. 6.0) in phosphate and MES buffers, similar results could be obtained by using HEPES at pH 7.0. The pH optimum for product formation was also analyzed by surveying TKS activity in ADA buffer between pH 5.7 – 7.8 allowing for a better pH range using only one buffer. The maximum production of tetraketides (HTAL and olivetol) occurred at a pH 6.4, whereas maximum PDAL production was at pH 7.3 (Figure 3.7). No OA production was observed.

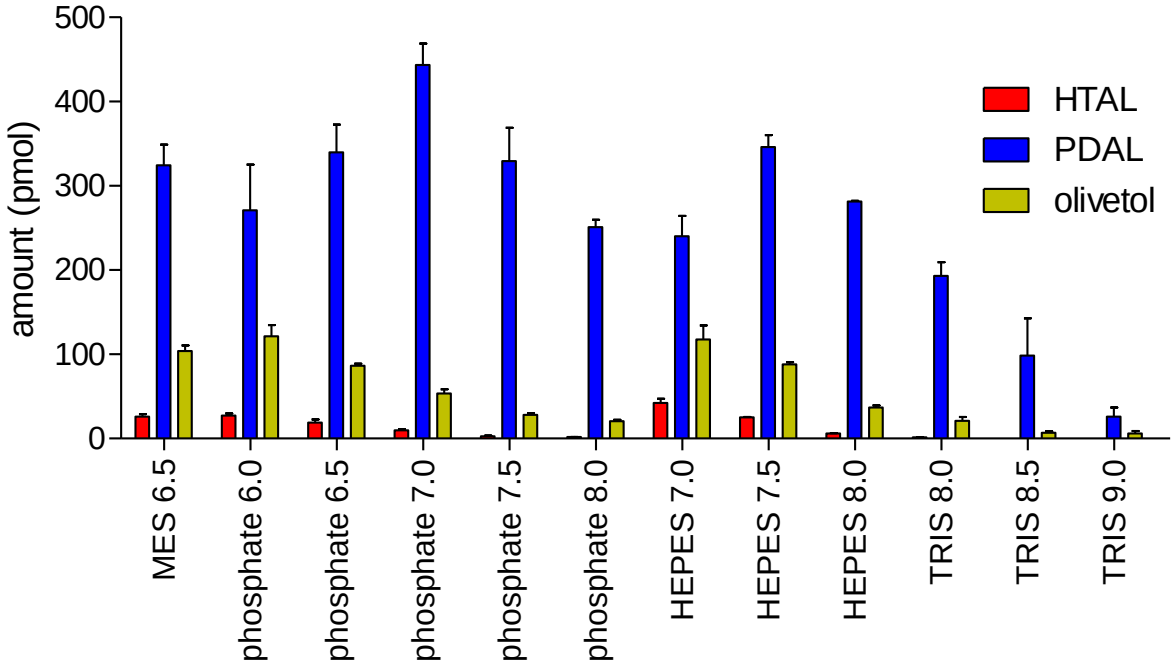


Figure 3.6 TKS products in different pH and buffers.

TKS product formation was first assessed in a pH range of 6.5 – 9.0 using a variety of buffers. HEPES [pH 7.0] was determined to be the most stable buffer with respect to TKS activity based on this data set. Error bars represent SD where n=3.

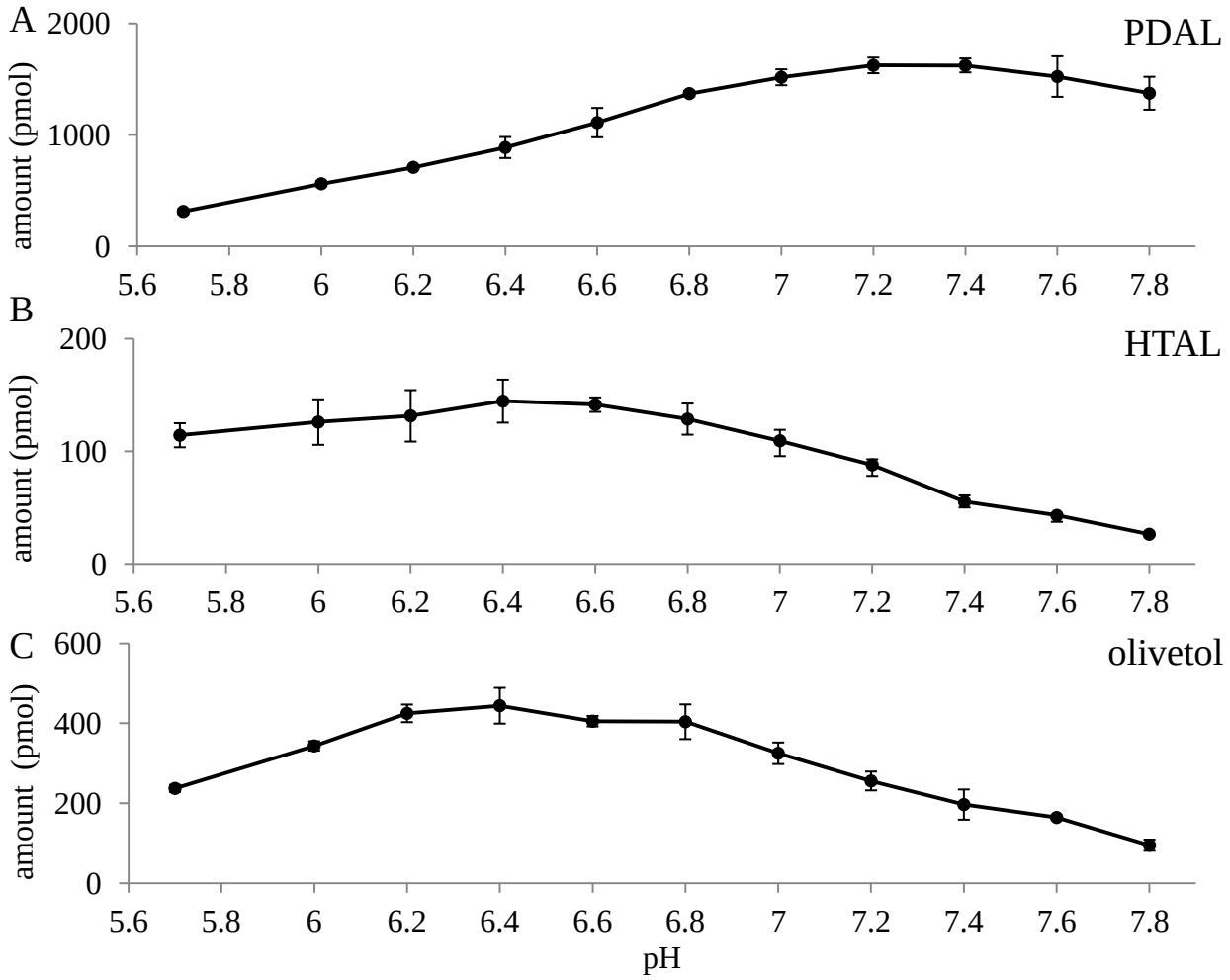


Figure 3.7 TKS product formation in ADA buffer.

The accumulation of TKS products was monitored between pH 5.7 and 7.8 in 20 mM ADA buffer. Product analysis for (A) HTAL, (B) PDAL, and (C) olivetol are shown in the corresponding panel. Error bars represent SD where n=3.

3.3.3.3 Temperature

To determine the effect temperature had on TKS activity, assays were performed between 10°C and 50°C in 100 mM HEPES buffer at pH 7.5. This pH was chosen since it corresponded to the maximum tetraketide production. Olivetol production appeared static until 40°C, but decreased at 50°C (Figure 3.8). Comparatively, HTAL and PDAL synthase activity appeared to increase until 40°C. PDAL production was slightly reduced 50°C, whereas HTAL production was approximately the same as seen in assays conducted at 40°C.

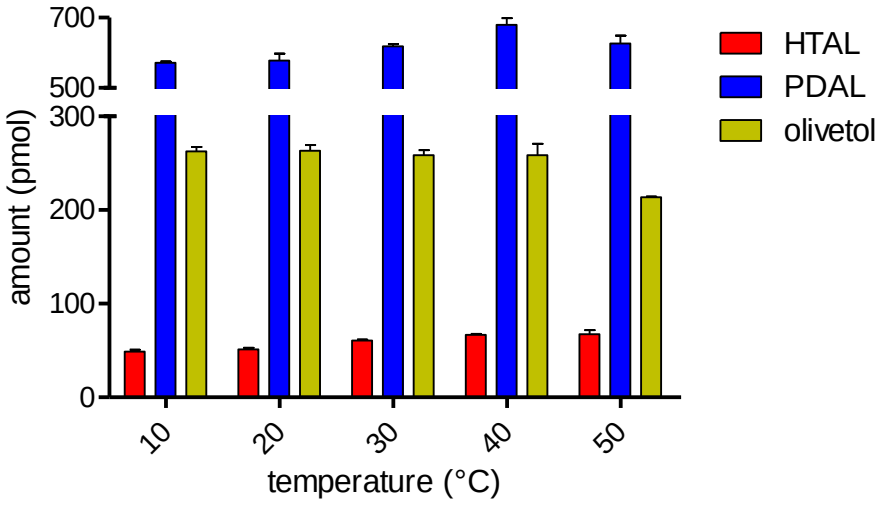


Figure 3.8 Temperature affects TKS product formation.

The accumulation of TKS products HTAL (red), PDAL (blue) and olivetol (green) was determined between 10 to 50°C. Assays were conducted in 100 mM HEPES [pH 7.5] in the presence of 0.2 mM malonyl-CoA and 0.2 mM hexanoyl-CoA. Vertical axis is split to show smaller amounts of product accumulation. Error bars represent the SD where n=3.

3.3.3.4 Substrate Availabilty

TKS was also assayed in the presence of 0.10 to 1.0 mM malonyl-CoA while 0.2 mM hexanoyl-CoA was maintained, and similarly in the presence of 0.10 – 1.0 mM hexanoyl-CoA while 0.2 mM malonyl-CoA was maintained. The concentration of both malonyl-CoA and hexanoyl-CoA affected the chemoselectivity of the TKS reaction. No product was detected when 0.10 mM malonyl-CoA was used in assays, but an increase in malonyl-CoA concentration correlated with an increase in product formation, except for PDAL which slightly decreased when 1.0 mM malonyl-CoA was used (Figure 3.9). Maximum PDAL production (938 ± 123 pmol) was detected when TKS was assayed with of 0.5 mM malonyl-CoA and 0.2 mM hexanoyl-CoA. Both HTAL (54 ± 8.4 pmol) and olivetol (79.0 ± 26.5 pmol) production were increased when TKS was assayed with 1.0 mM malonyl-CoA.

The availability of hexanoyl-CoA also affected TKS product formation (Figure 3.10). Olivetol accumulated to greater amounts as hexanoyl-CoA increased, whereas HTAL and PDAL production peaked at 0.2 mM hexanoyl-CoA.

Product ratios used to assess TKS triketide:tetraketide chemoselectivity indicate PDAL is preferentially formed when TKS was assayed with 0.1 mM hexanoyl-CoA and 0.2 mM malonyl-CoA (Table 3.4). Conversely, increasing either malonyl-CoA or hexanoyl-CoA concentrations to 1.0 mM resulted in the increased synthesis of tetraketides. Product ratios were also used to evaluate HTAL:olivetol chemoselectivity, which clearly reflect the preferential production of olivetol when hexanoyl-CoA concentrations were raised (Table 3.4). The triketide:tetraketide ratio represents total PDAL divided by the sum of HTAL and olivetol.

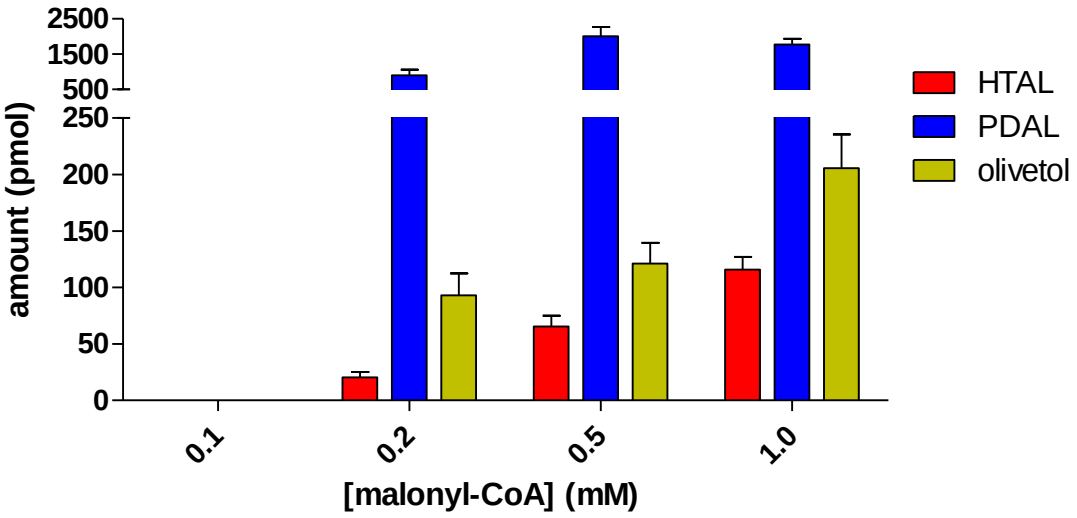


Figure 3.9 TKS product formation in varying malonyl-CoA concentrations.

The amount of TKS products HTAL (red), PDAL (blue), and olivetol (green) were determined after TKS was assayed in the presence of 0.2 mM hexanoyl-CoA and 0.1 to 1.0 mM malonyl-CoA. Vertical axis is split to show smaller amounts of product. Error bars are the SD where n=3.

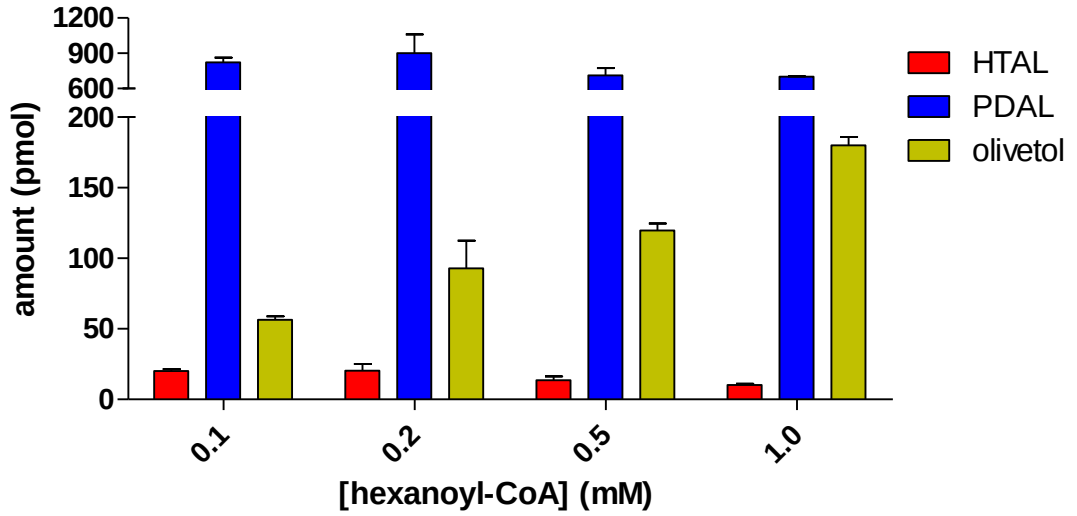


Figure 3.10 TKS product formation in varying hexanoyl-CoA concentrations.

The amount of TKS products HTAL (red), PDAL (blue), and olivetol (green) were determined after TKS was assayed in the presence of 0.2 mM malonyl-CoA and 0.1 to 1.0 mM hexanoyl-CoA. Vertical axis is split to show smaller amounts of product. Error bars are the SD where n=3.

Table 3.4 Ratios of product formed during TKS assays.

	triketide : tetraketide*	HTAL : olivetol
0.1 mM malonyl-CoA	nd	nd
0.2 mM malonyl-CoA	11.4	0.31
0.5 mM malonyl-CoA	11.6	0.59
1.0 mM malonyl-CoA	6.2	0.68
0.1 mM hexanoyl-CoA	13.2	0.42
0.2 mM hexanoyl-CoA	11.4	0.31
0.5 mM hexanoyl-CoA	9.3	0.20
1.0 mM hexanoyl-CoA	7.1	0.11

* Ratios for triketide/tetraketide were determined by comparing the amount of PDAL by the sum amount of HTAL and olivetol derived from Figure 3.9 and 3.10.

nd = not detected.

3.3.3.5 TKS Activity in the Presence of DTT

The use of dithiothreitol (DTT) and β-mercaptoethanol (βME) during protein purification, as well as the use of DTT in TKS assays, required an analysis of the potential deleterious effects caused by these strong reducing agents. Two solutions of TKS derived from the same culture were purified in tandem; one in the presence of reducing agents (DTT and βME), the other in their absence. Both protein solutions (± DTT/ βME) were assayed in increasing concentrations of DTT and product formation was monitored. TKS purified in the absence of DTT and βME possessed < 70% activity of TKS purified in the presence of DTT and βME (Figure 3.11). The addition of 5 mM DTT to TKS assays increased HTAL and PDAL production by 69% and 56%, respectively, although olivetol formation decreased by 10%. TKS activity was reduced when DTT concentrations exceeded 5 mM, and resulted in the least amount of product when TKS was assayed in the presence of 100 mM DTT.

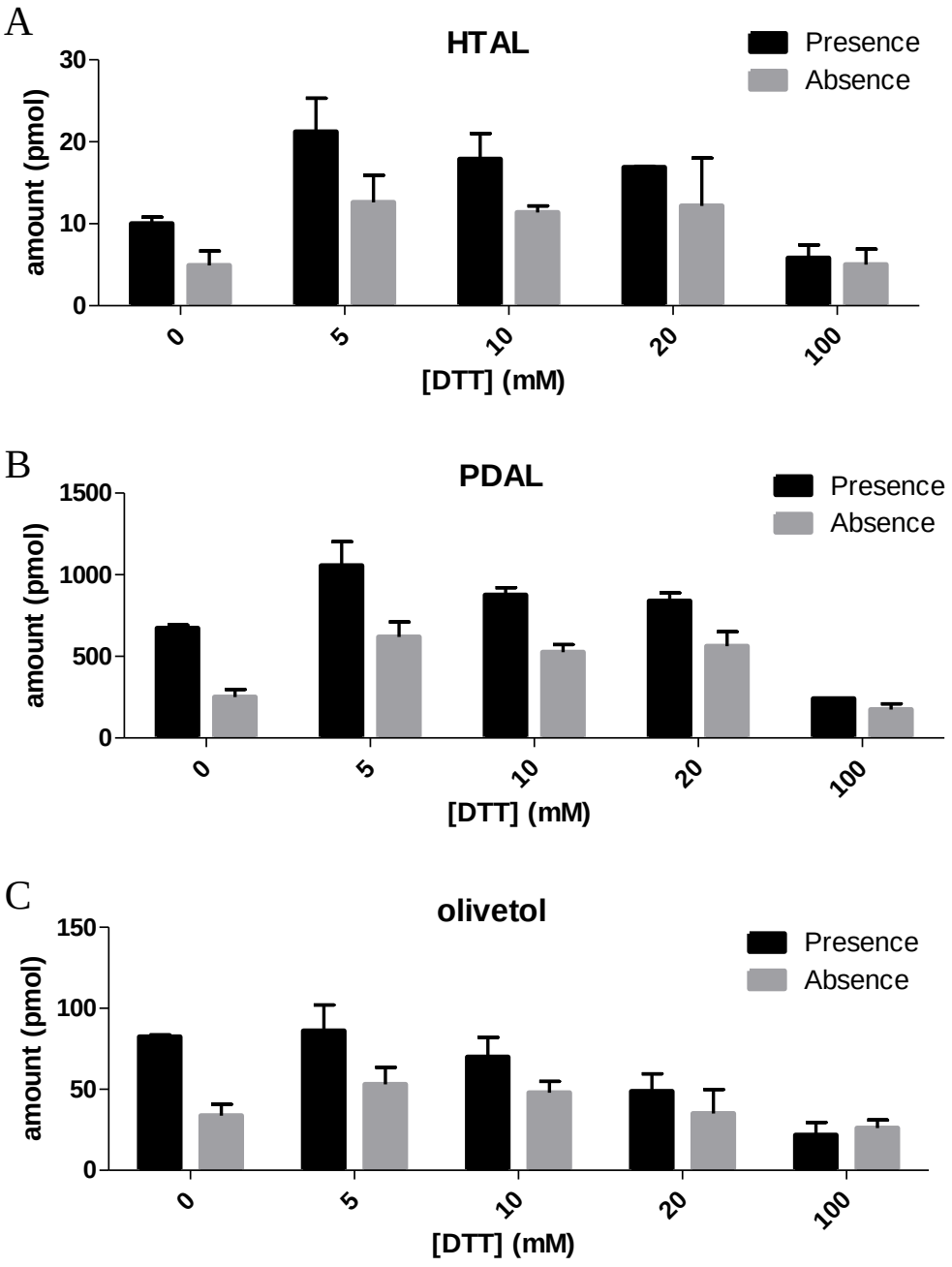


Figure 3.11 TKS in the presence and absence of DTT and βME during purification.

Two different preparations of recombinant TKS were assayed; one purified with DTT and βME (black), the other purified without DTT and βME (grey). Assays were conducted in increasing concentrations of DTT (0 to 100 mM) to determine if TKS product formation would be affected. Panels show the production of (A) HTAL, (B) PDAL, and (C) olivetol. Error bars are the SD where n=3.

3.3.4 OA Production by Trichome Protein Extracts

Protein extracts from isolated trichomes obtained from drug-type and USO-31 hemp plants were separately assayed. Trace amounts of OA were detected in both samples by HPLC analysis implying OA synthase activity (Figure 3.12). Whereas USO-31 hemp trichome extracts accumulated HTAL, PDAL, OA and olivetol, no olivetol was detected in drug-type trichome extracts. In an attempt to increase the synthesis of assay products, trichome protein extracts obtained from USO-31 hemp were assayed in a scaled-up reaction using ten times more the starting materials than previous assays (Figure 3.12). In this larger reaction, HTAL, PDAL, OA, and olivetol accumulated to a ratio of 1:4:1:2, respectively. Absolute amounts were determined to be 167 pmol HTAL, 658 pmol PDAL, 161 pmol OA, and 318 pmol of olivetol.

Drug-type trichome activity was further investigated in the presence and absence of recombinant TKS. Reactions used the same amount of TKS and/or trichome extracts in each assay. Trichome extracts, when assayed in the presence of TKS, produced 638% more OA than trichome extracts alone (Figure 3.13). HTAL and olivetol yields were also greater in the combined assays than in assays performed using only trichome protein (858% and 616%, respectively) or recombinant TKS (209% and 180%, respectively). The ratio of triketides/tetraketides formed was also markedly different when TKS was assayed alone (ratio 8:1), than when trichome extracts were assayed in the absence or presence of TKS (ratio 1:4 and 1:0, respectively). The HTAL:PDAL:OA:olivetol product ratio for: recombinant TKS assays was 1:45:0:4; trichome protein extracts was 1:7:2:5; for trichome protein extracts in the presence of recombinant TKS was 1:9:2:4.

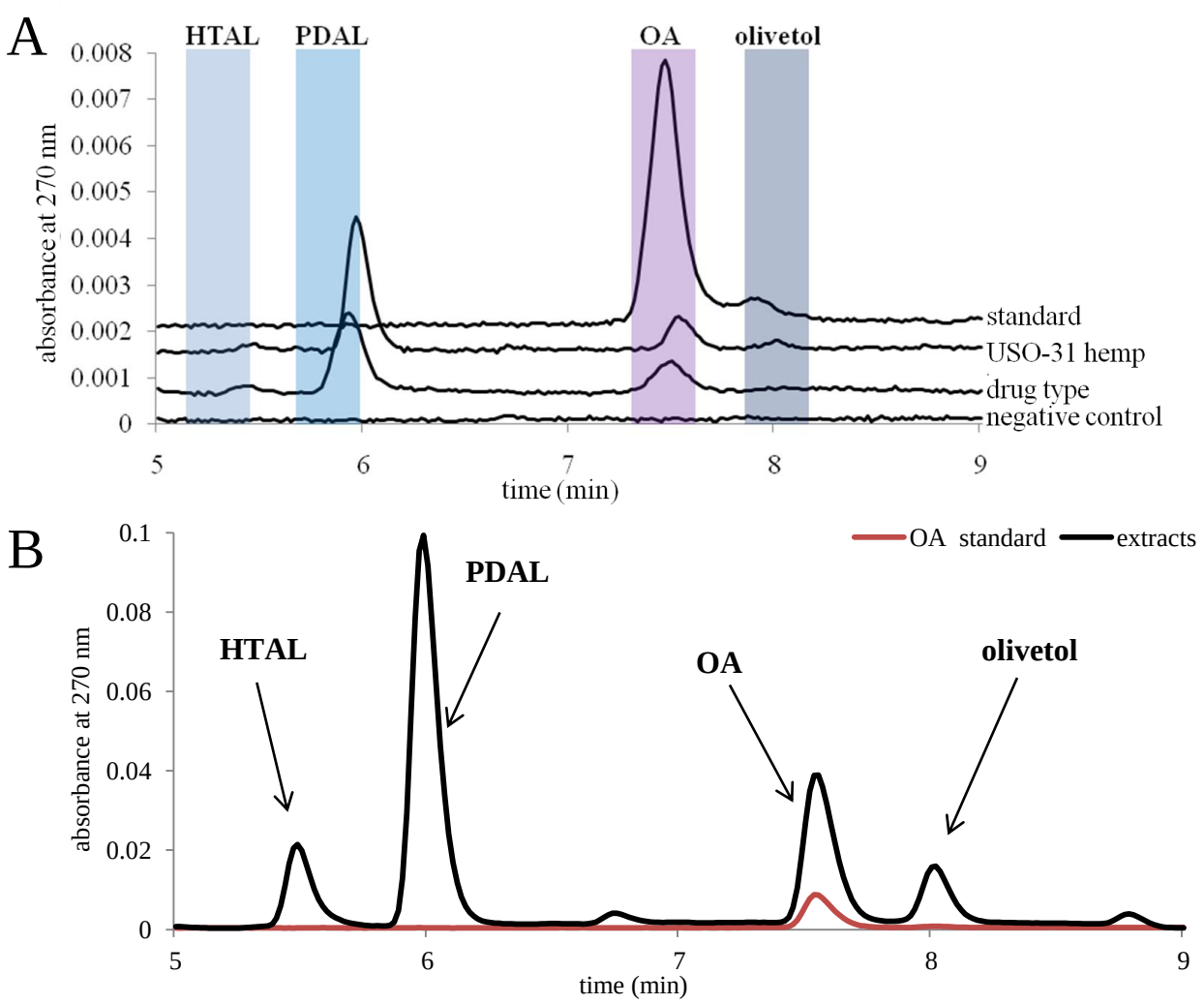


Figure 3.12 Assays of cannabis trichome protein extracts

(A) HPLC chromatograms of products formed during cannabis trichome extract assays indicate the accumulation of OA when compared to an OA standard, except for trichomes assayed in the absence of malonyl-CoA (negative control). (B) An LC chromatogram of products formed during a larger USO-31 hemp trichome extract assay clearly detected HTAL, PDAL, olivetol, and OA.

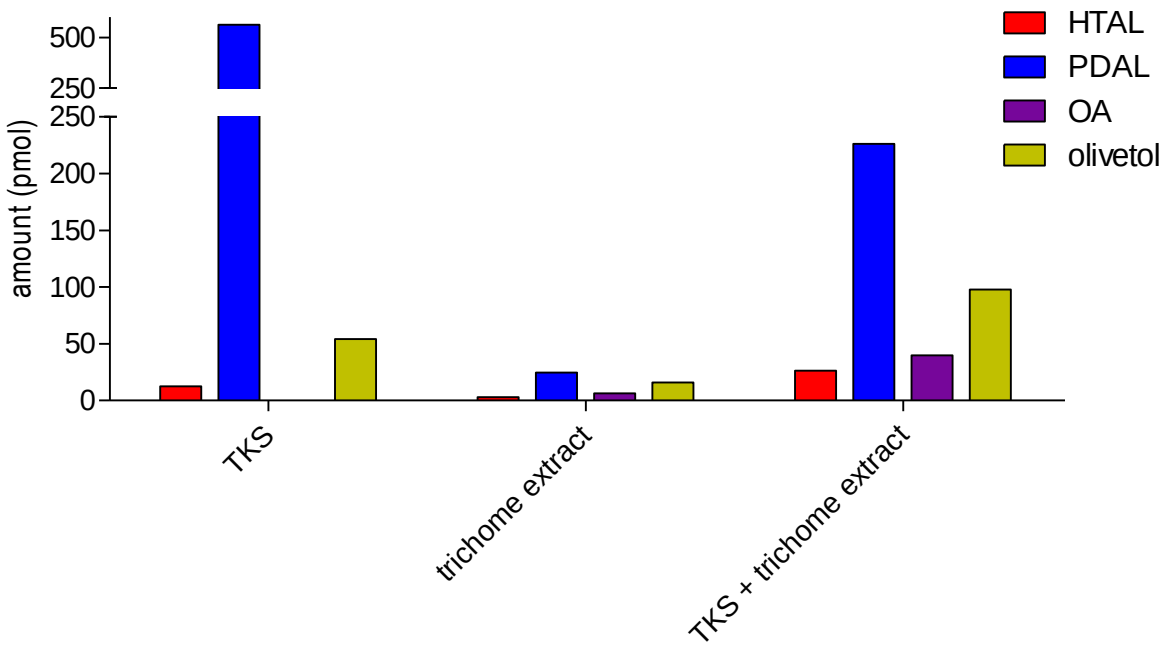


Figure 3.13 Drug-type trichome extracts assayed with recombinant TKS.

Product yields obtained when TKS and trichome extracts were assayed together or alone indicate the amount of HTAL (red), PDAL (blue), OA (purple), and olivetol (green) detected. Vertical axis is split to show smaller amounts of product accumulation.

3.3.4.1 Ammonium Sulfate Precipitation of Drug-Type Trichome Protein Extracts

Once it was determined that trichomes possess OA synthase activity (Figure 3.12), ammonium sulfate was used to precipitate trichome proteins, which might have been able to separate olivetol and OA synthase activity, thus implying that separate proteins are required during OA biosynthesis. A protein solution obtained from isolated trichomes was precipitated using increasing amounts of ammonium sulfate (20%, 40%, 60%, and 80% w/v) followed by centrifugation. The precipitates were resuspended in 100 mM HEPES buffer [pH 7.5]. An aliquot from each fraction, along with a sample of the remaining supernatant was separated by electrophoresis before being transferred to a PVDF membrane (Figure 3.14A and 3.14B). A Western blot was then performed using polyclonal αTKS antibodies, and showed that the greatest accumulation of TKS occurred in the 80% cut (Figure 3.14). Polyclonal αTKS antibodies also detected impurities in recombinant TKS samples, and to a lesser degree, in each of the precipitated fractions.

The precipitated protein fractions, along with the remaining unprecipitated supernatant and a positive TKS control, were separately assayed in the presence of hexanoyl-CoA and malonyl-CoA to determine if any of the precipitated protein fractions retained OA synthase activity. Resuspended protein derived from the 80% salt cut had OA synthase activity, but other fractions, including the supernatant, had neither TKS nor OA synthase activity, except for what appeared to be a trace amount of OA detected in the 0% fraction (Figure 3.14).

Due to the lack of PDAL, HTAL, and olivetol, the appearance of what appears to be OA in 0% ammonium sulfate fractions is believed to be a contaminant.

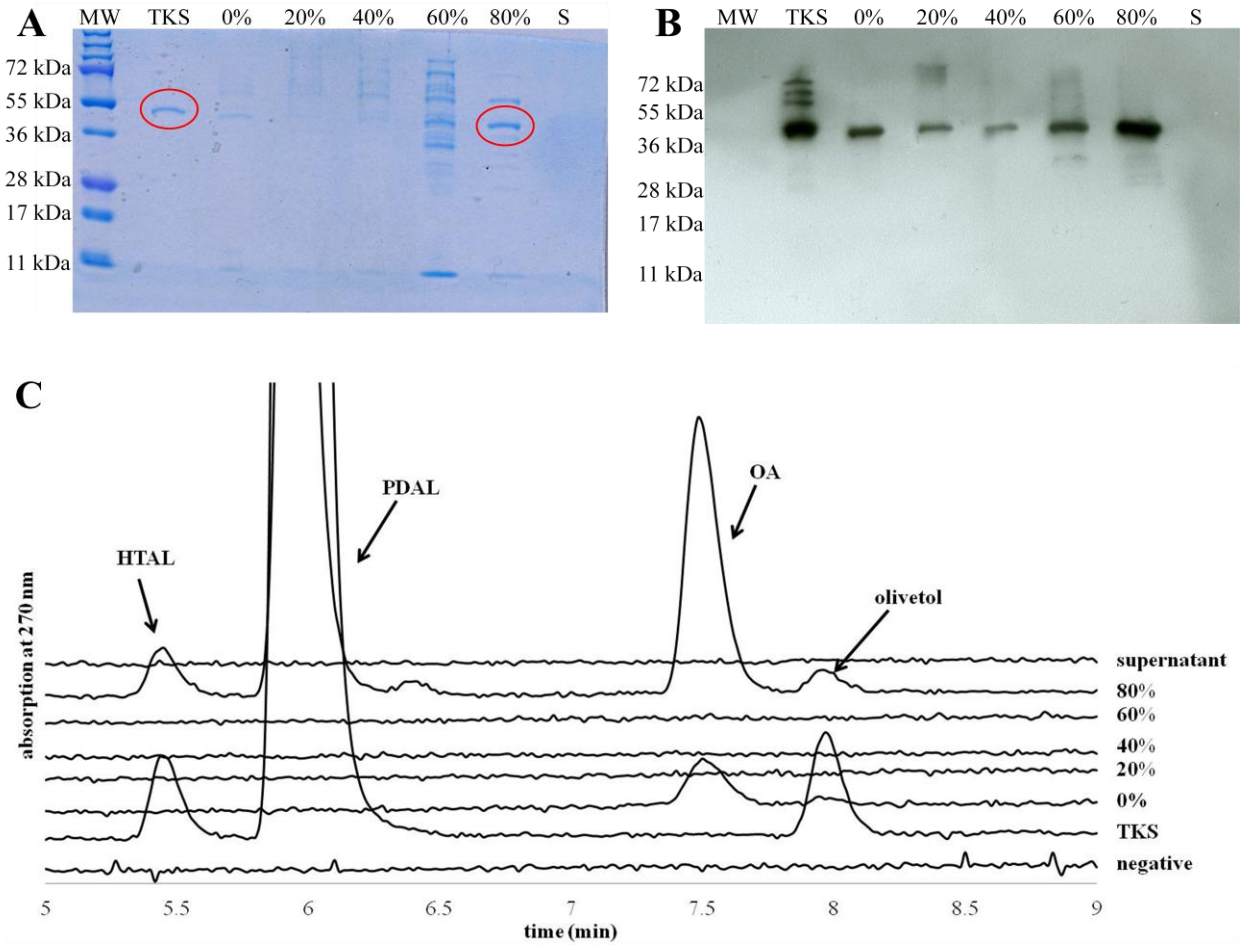


Figure 3.14 Activity of trichome protein extracts in ammonium sulfate.

(A) 10% SDS-PAGE of equal volume of trichome protein precipitates occurring at 0%, 20%, 40%, 60%, and 80% ammonium sulfate fractions were run with Protein Ladder (MW), 1 µg of recombinant TKS (TKS), and an aliquot of unprecipitated supernatant obtained from post 80% ammonium sulfate treatments (S). TKS is circled in red. (B) A Western blot obtained after transferring proteins from (A) to a PVDF membrane indicates the majority of TKS is detected in 80% ammonium sulfate fractions. (C) HPLC traces of assay products produced from recombinant TKS and trichome extracts obtained from the ammonium sulfate precipitates are offset for comparison.

3.3.4.2 The Co-Immunoprecipitation of Trichome Protein Extracts

Polyclonal αTKS antibodies were used in co-immunoprecipitation experiments designed to retrieve TKS and potentially interacting proteins from drug-type trichome protein extracts. Purified rabbit IgG anti-TKS antibody-linked resin were incubated with either 1) a crude recombinant TKS solution, or 2) a protein solution extracted from isolated trichomes. Control resin was also used to determine non-specific protein binding to the resin. The antibody-linked resin was capable of successfully binding TKS from solutions, as seen when SDS-PAGE with Coomassie stain was used to visualize proteins (Figure 3.15). Another protein of (ca.) 60 kDa was also purified from both recombinant and trichome extracts, and sent to the PBI mass spectrometry lab for Q-TOF MS/MS analysis. The unknown protein(s) was identified as an *A. thaliana* chaperonin homolog (annotated NP_175945) (Appendix C). Chaperonins are highly conserved proteins involved in the proper folding of other proteins (Walter and Buchner, 2002). The presence of TKS was also confirmed by Q-TOF MS/MS analysis (Appendix D).

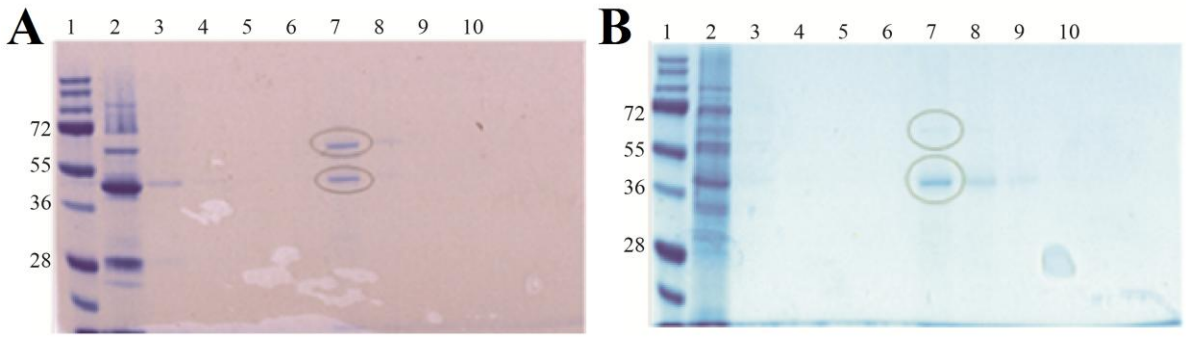


Figure 3.15 Co-immunoprecipitation of TKS and trichome protein extracts

Co-immunoprecipitation was conducted on (A) recombinant TKS and (B) trichome protein extracts using polyclonal αTKS antibody-coupled resin. The resin was capable of binding recombinant TKS (A; lane 7; 45 kDa) and an unknown protein (A; lane 7; ca. 60 kDa), as well as TKS found in trichome extracts (B; lane 7; 42 kDa) and another protein (B; lane 7; ca. 60 KDa). Lanes contained (1) Protein Ladder (numbered in kDa), and 20 µL of (2) flowthrough from resin columns, (3) first wash, (4) second wash, (5) third wash, (6) fourth wash, (7) first elution, (8) second elution, and (9) third elution. Drawn circles show the eluted proteins.

3.3.4.3 Protein Pull-Down Assays of Trichome Protein Extracts

Recombinant TKS was bound to Talon beads and incubated with trichome protein extracts in an attempt to capture proteins that interact with TKS. Results from this experiment did indicate the presence of a ca. 30 kDa protein (Figure 3.16A; lanes 3 and 4) that was not present in a fresh TKS control (Figure 3.16A; lane 2). Although dilute, trichome extracts did not show this band (Figure 3.16A; lane 5 and 6). The unknown protein is assumed to be degraded and fragmented TKS as it was also detected in older TKS-only preparations (Figure 3.16B).

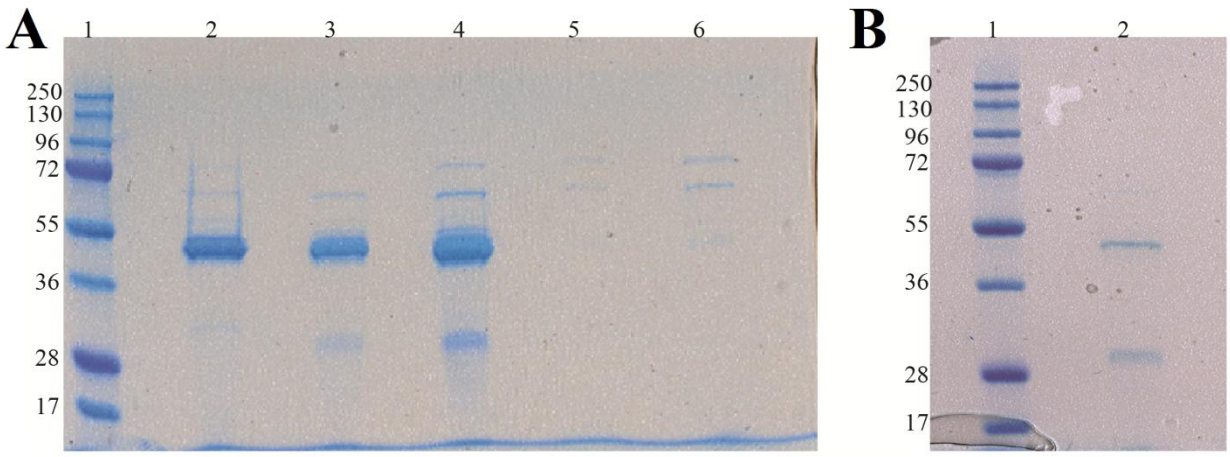


Figure 3.16 Protein pulldown using TKS and cannabis drug-type trichome extracts.

(A) SDS-PAGE of proteins used in an attempt to pull down potentially interacting proteins from trichome extracts. Lanes contained (1) 10 µL Protein ladder (numbered in kDa), (2) 2 µg recombinant TKS, (3) 3 µg recombinant TKS following incubation with trichome extracts, (4) 6 µg recombinant TKS following incubation with trichome extracts, (5) 1 µg trichome extract, (6) 2 µg trichome extract. (B) A 10% acrylamide gel containing (1) 10 µL Fermentas Prestained Protein ladder, and (2) 1 µg recombinant TKS. Although a new band appears in (A) lanes 3 and 4 at ca. 30 kDa, this was presumed to be a TKS degradation product after it was also observed in (B) a TKS-only sample electrophoresed two days later.

3.3.5 The Analysis of cDNA Cyclase Candidates

The goal of experiments was to identify, clone and express trichome-specific proteins that were potentially involved in the synthesis of OA. Candidate transcripts that were both highly abundant in the cDNA library and showed similarity to proteins with cyclase or cyclase-like activity were sub-cloned, expressed and purified before being assayed in the presence and absence of TKS (Figure 3.17). Candidate proteins included a putative chalcone isomerase (CHI; annotated Q6V7U9), a Betv1-like protein (pathogenesis related protein PR10; annotated B5KVN9) and a DABB protein (probable protein Pop3; annotated Q9LUV2) (Table 3.1). The inclusion of CHI was based on it constituting 0.52% of the total transcripts in trichome ESTs and its relationship with CHS in other plants, where CHI is capable of cyclizing chalcone to chalcone naringenin (Gensheimer and Mushegian, 2004; Jez et al., 2001). The Betv1-like protein (also known as major latex protein/MLP) constituted 0.25% of the total trichome ESTs, and shares structural similarity to *Streptomyces* TcmN ARO/CYC, a PKC involved in the biosynthesis of tetracenomycin (Ames et al., 2008). The DABB protein is also well represented in the trichome ESTs, representing 0.45% of the total library, and is structurally similar to *Streptomyces* TcmI, a cyclase also involved in the biosynthesis of tetracenomycin (Thompson et al., 2004).

Candidate proteins were assayed in the presence and absence of TKS to determine if they produced OA at a ratio of 100 pmol TKS (ca. 2 µg) to 500 pmol of candidate protein (ca. 20 µg). Assays were performed in triplicate and their products analyzed by HPLC (Figure 3.18). HPLC chromatograms clearly show that, when TKS is assayed in the presence of the DABB candidate cyclase, OA is detectable, along with HTAL, PDAL, and olivetol. No OA accumulated when TKS was assayed alone, or in the presence of MLP or CHI, although the other TKS products (*i.e.*

PDAL, HTAL, and olivetol) were detected. The accumulation of TKS products also varied between samples (Figure 3.19). The amount of PDAL appeared invariable amongst the samples tested, with an average of 2357 ±SE 45.6 pmol produced (where n = 3). The amount of HTAL produced appeared static when TKS was assayed alone, or with the DABB protein, or CHI, with an average of 193 ±SE 4.3 pmol; less was produced when TKS was assayed in the presence of MLP (119 ±SD 10 pmol). The production of olivetol was the most varied between samples, with an average of 389 ±SE 20 pmol when TKS was assayed alone or in the presence of CHI, but less when TKS was assayed with the DABB protein or with MLP, producing 166 ±SD 32 pmol and 228 ±SD 27 pmol of olivetol respectively. The ratio of HTAL to olivetol produced when TKS was assayed in the presence of the DABB protein is 1.2, whereas it was 0.5 when TKS was assayed in the presence or absence of CHI or MLP.

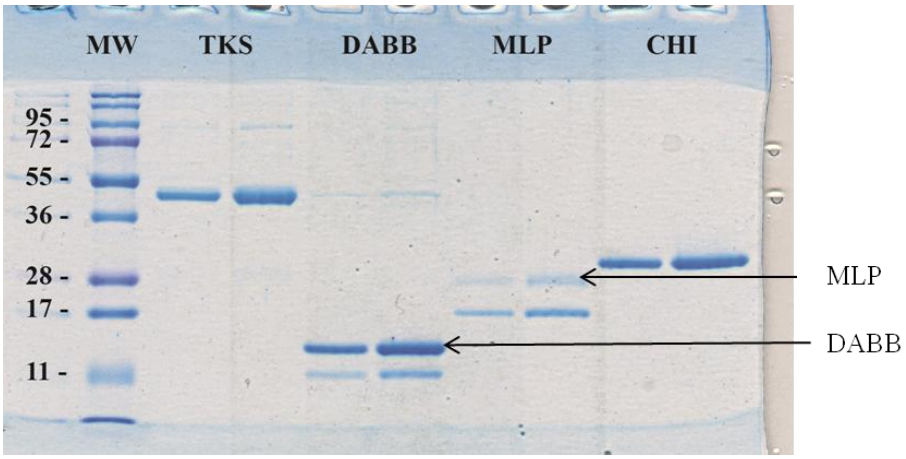


Figure 3.17 SDS-PAGE of TKS and candidate cyclases.

SDS-PAGE was used to analyze recombinant TKS (TKS; 44.5 kDa), and putative tetraketide cyclases including a DABB protein (18.7 kDa), Betv1-like protein (MLP; 20.7 kDa), and a chalcone isomerase-like protein (CHI; 27.6 kDa). Each IMAC-purified protein was loaded onto the 12% acrylamide gel using (first lane) 1.25 µg and (second lane) 3.75 µg of protein and were run next to a molecular weight marker (MW).

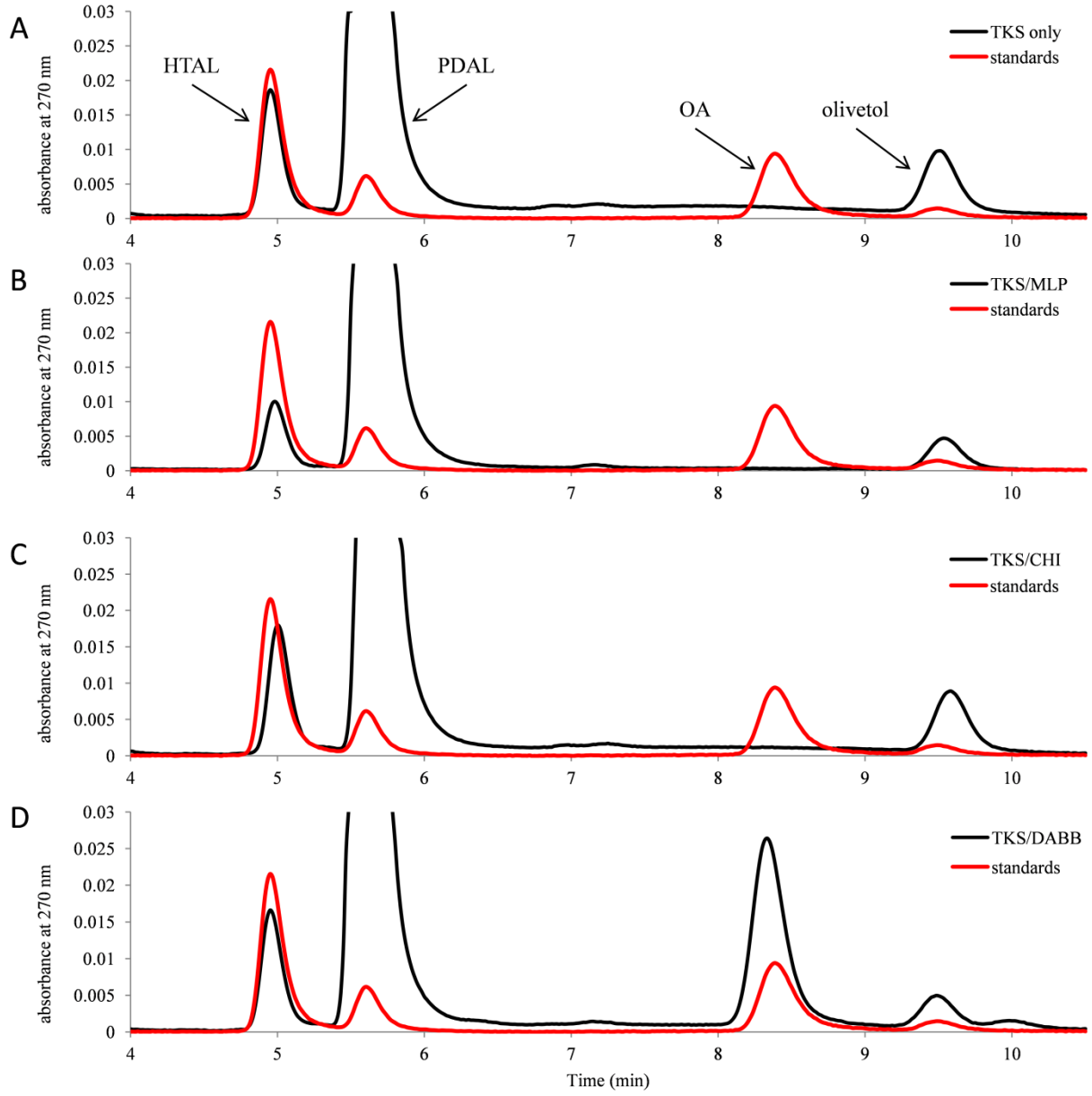


Figure 3.18 TKS activity in the presence of candidate cyclases.

HPLC traces of the products formed when TKS was assayed (A) alone, or with (B) MLP-like, (C) CHI-like, or (D) DABB proteins are overlayed. Whereas HTAL (RT ca. 5.0 min), PDAL (RT ca. 5.5 min), and olivetol (RT ca. 9.5 min) were detected in all assays, OA (RT ca. 8.3 min) was only detected when TKS was assayed in the presence of the DABB protein. Standards used during HPLC analysis are overlayed (red trace).

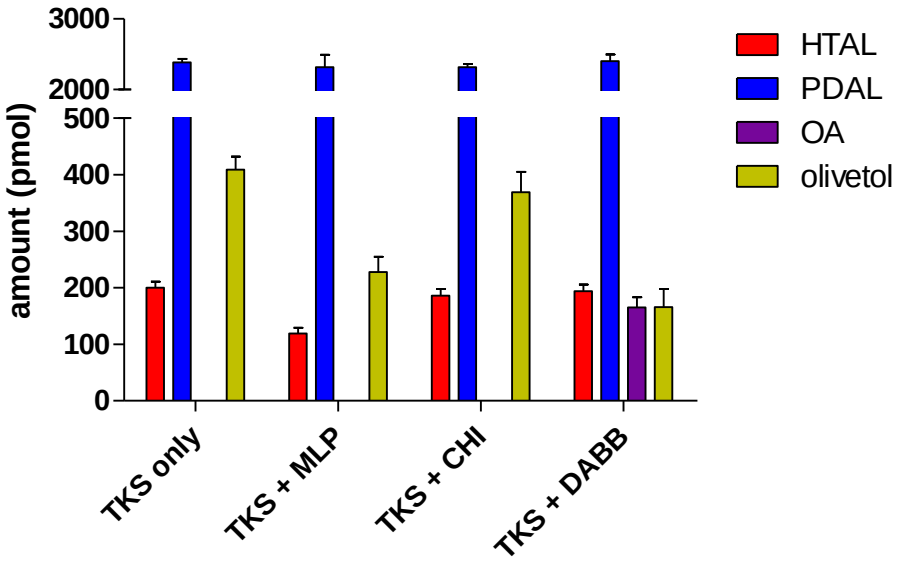


Figure 3.19 TKS assays products when assayed in with putative cyclases.

The absolute pmol amount of TKS products detected when TKS was assayed in the presence and absence of CHI, MLP, and DABB proteins is shown. OA was only detected when TKS was assayed in the presence of the DABB candidate. Vertical axis is split to show smaller amounts of product accumulation. Error bars are the SD where n=3.

3.4 Discussion

Cannabis contains many secondary metabolites, including cannabinoids, stilbenes, anthocyanins, flavonoids, and spirans, that are synthesized by different PKSs (Flores-Sanchez and Verpoorte, 2008b). The biosynthesis of these compounds requires type III PKSs that accept different starting substrates and vary in their chain lengthening activity and cyclization modes. The first cannabis CHS reported in the literature was identified using degenerate primers based on the sequences of the *H. lupulus* CHS and phlorisovalerophenone synthase (VPS). This cannabis CHS was described as a CHS and VPS (Raharjo et al., 2004b). It is capable of accepting p-coumaroyl-CoA as a reaction initiator, but fails to use hexanoyl-CoA as a substrate, making it incapable of producing OA. The subsequent identification of different PKSs in cannabis revealed an evolutionary divergence between PKSs that cluster either with known CHSs, such as the *H. lupulus* CHS/VPS (annotated CAC19808), or with CHS-like sequences, which are not involved with the biosynthesis of naringenin chalcone (Flores-Sanchez et al., 2010). TKS shares 100% amino acid identity with PKSG4, a CHS-like paralog that has only been characterized using *in silico* methods (Flores-Sanchez et al., 2010). TKS also shares 98.7% amino acid identity with another CHS-like homolog, OLS, which has been biochemically characterized and has had its crystal structure solved (Taguchi et al., 2008; Taura et al., 2009). Although the characterization of OLS has been conducted, the discovery of TKS occurred at approximately the same time. The partial characterization of TKS which is described in this chapter indicates that TKS has similar activity as OLS. Amino acid differences between OLS and TKS do not appear to affect olivetol synthase activity, although a future study could investigate all known TKS homologs for further robust analysis.

The characterization of TKS in this study was not conducted under the same conditions as used for OLS (Taura et al., 2009). In the present work, a variety of buffers were used to determine the optimum for the production of tetraketides, and although TKS assays conducted in ADA buffer showed an increased tetraketide production at pH 6.4, a greater absolute amount of tetraketides, as well as a lower triketide:tetraketide ratio resulted when TKS was assayed in HEPES buffer at pH 7.0, and therefore further assays were conducted using HEPES. The observed changes in TKS activity when assayed in different buffers may be attributed to the enzyme's conformational stability (Ugwu and Apte, 2004). All *in vitro* assays resulted in PDAL and HTAL production, which are suspected to be early derailment products resulting from artificial *in vitro* conditions (Austin and Noel, 2003). It is not known if the biosynthesis of pyrones by TKS has a specific biological function, or if they are even produced in the plant, although the expression of *Arabidopsis* CHS-like LAP5 and LAP6, which also produce alkyl pyrones during *in vitro* assays, have been shown to be required for pollen exine development which plays a crucial role in plant reproduction (Dobritsa et al., 2010). Early characterization work continued using HEPES buffer at pH 7.5, since a greater amount of tetraketides were produced. Contrastingly, Taura et al. (2009) assayed OLS in citrate buffer at a pH 5.5. TKS and OAC assays with citrate buffers are shown in Chapter 4.

The temperature at which assays were conducted in earlier work was determined to be optimum at 40°C based on the overall amount of products formed (Figure 3.8). At 50°C, both olivetol and PDAL accumulated to a lesser degree than at lower temperatures, whereas the production of HTAL was slightly increased. An increase in HTAL production, coupled to a decreased accumulation of PDAL and olivetol, suggest that the overall enzyme activity is

decreased at > 40°C. Furthermore, the increase of HTAL at higher temperatures coupled to a loss of olivetol suggests that HTAL and olivetol share a common tetraketide intermediate pre-cursor. The lactonization of a tetraketide intermediate to HTAL may be a spontaneous and thermodynamically-favored process compared to olivetol cyclization, which may require a slower step so to achieve cyclization. Alternatively, enzyme stability is compromised at temperatures > 40°C, leading to an increase in enzyme derailment products.

Substrate availability assays indicate that increased malonyl-CoA concentrations result in a lower triketide:tetraketide ratio (Figure 3.9 and Table 3.4). This intuitively seems correct, since a growing polyketide chain that is attached to a PKS active site requires malonyl-CoA to produce a final, longer product, and that a (shorter) derailment product (*i.e.* triketide leading to PDAL) would result if less malonyl-CoA was available. Although a greater malonyl-CoA concentration favors the production of tetraketides, HTAL production is also favored over the production of olivetol under these conditions. An increase in the availability of hexanoyl-CoA also decreased the triketide:tetraketide product ratio, but a greater amount of olivetol was synthesized under these conditions instead of HTAL (Figure 3.10 and Table 3.4). These results show that TKS chemoselectivity can be altered by varying substrate concentration and suggest that excess malonyl-CoA, or possibly its breakdown product, acetyl-CoA, either destabilize an enzyme-tethered tetraketide intermediate causing its early release and cyclization to HTAL, or influences the cyclization of a linear tetraketide intermediate that has already been released by TKS, to favor HTAL production. The use of a malonyl-CoA:hexanoyl-CoA substrate ratio of 3:1 in subsequent assays was used based on the stoichiometry of tetraketide biosynthesis.

DTT is a strong reducing agent routinely added to buffers used during the purification and assaying of enzymes. DTT is used to prevent cysteine residues from forming disulfide linkages within and between proteins, and to assure the reduction of the cysteine residue in the TKS active site. Although DTT has been used extensively, there was concern regarding the potential interfering effects which could present themselves as a problem in TKS reactions. Type III PKS activity is dependent on the modification of thiol bonds involving CoA esters (Austin and Noel, 2003). Since a similar reaction mechanism is involved between type III PKSs and DTT through the use of sulfhydryl bonds, the additive and deleterious effects of DTT were examined in TKS assays. Results indicate that the addition of DTT and βME in purification buffers has a beneficial effect on TKS activity, although the use of DTT in enzyme assays increases the production of pyrones, but not olivetol, at lower concentrations, and negatively affects all TKS activity at concentrations > 5 mM. The use of DTT in enzyme assays was deemed beneficial at concentrations of 5 mM.

The assaying of trichome-derived protein solutions provided direct evidence of OA synthase activity in cannabis (Figure 3.12). Earlier studies that attempted to detect OAS activity from cannabis tissues were unable to detect OA, and reported the production of olivetol instead (Flores-Sanchez and Verpoorte, 2008a; Raharjo et al., 2004a). In this study, when a USO-31 hemp trichome protein solution was assayed, HTAL, PDAL, OA and olivetol were detected, whereas only HTAL, PDAL, and OA were detected when a drug-type trichome protein solution was assayed. The small amount of drug-type trichomes that were used may have not been enough for the production of olivetol in detectable amounts. The availability of drug-type trichomes restricted further analysis. Protein extract assays strongly suggest that TKS is involved

in the biosynthesis of OA since the addition of TKS to protein extract assays resulted in a 5-fold increase of OA (Figure 3.13). Assays confirmed that cannabis trichomes possess OA synthase activity, and led to further attempts at identifying other proteins required for the proper cyclization of tetraketide products. The analysis of trichome extracts following treatments with ammonium sulfate show that the OA synthase activity is retained in the 80% ammonium sulfate precipitate, indicating that a protein and not a co-factor is involved. Furthermore, the isolation of the OA synthase activity along with the majority of TKS in the same fraction suggested the possibility that a physical interaction existed with TKS and the unknown protein. This instigated a search for potential TKS interaction partners using co-immunoprecipitation and protein pull-down methods.

Co-immunoprecipitation experiments used αTKS polyclonal antibodies that were bound to a solid-support agarose matrix. The finding of a chaperonin homolog was most likely due to the cross-reactivity of polyclonal antibodies that were generated using TKS contaminated with an *E.coli* chaperonin homolog (Figure 3.15). These experiments were unable to detect any protein that physically interacts with TKS. In another effort to discover a TKS-interacting protein, affinity chromatography (*i.e.* protein pull-down) was attempted after incubating trichome extracts in the presence of recombinant TKS, then purifying TKS and any interacting proteins using IMAC. An additional protein was detected during protein pull-down experiments, but it was concluded to be a TKS degradation product after a similar protein band was detected following the SDS-PAGE of TKS-only controls (Figure 3.16). Neither co-immunoprecipitation nor protein pull-down methods were either capable of retrieving any TKS-interacting proteins from trichome extracts, suggesting that no such protein exists, or that these methods were

incapable of detection. Experiments designed to isolate trichome proteins using chromatography had been planned, and attempts to tease out an accessory protein from trichome extracts were already underway, but were not continued since TKS assays conducted in the presence of potential cyclases meanwhile led to the discovery of OAC.

An analysis of the trichome-specific cDNA library resulted in the identification of several putative polyketide cyclases that could be involved in the biosynthesis of OA (Table 3.1). Another cannabis cDNA library prepared from isolated trichomes had also been published at the time of this study by Marks et al. (2009). TKS and its homolog (CAN24/OLS) in the Marks et al. (2009) library were highly expressed in EST libraries, where TKS is the most abundant transcript in the Stout et al. (2012) trichome cDNA library and CAN24 is the eighth most abundant transcript in the Marks et al. (2009) library. Similarly, the DABB protein (homologous to *A. thaliana* heat stable 1 protein; annotated NP_566569) chosen as a potential cyclase was also detected as a highly abundant transcript in both cDNA libraries, where it is the 15th-most abundant transcript in the Stout et al. (2012) trichome cDNA library and the fifth-highest in the Marks et al. (2009) library. The MLP-like and CHI-like transcripts were the 34th- and 9th-most abundant transcripts detected in the trichome cDNA library published by Stout et al. (2012) (Table 3.1), whereas they were the 10th- and 108th-most abundant in the Marks et al. (2009) library, respectively.

Assays containing TKS and the DABB candidate led to the production of OA, whereas TKS assays with the other candidate cyclases did not (Figure 3.18). Early TKS assays performed in the presence of DABB protein produced an equimolar amount of tetraketides, with an HTAL:PDAL:OA:olivetol ratio of approximately 1:10:1:1 (Figure 3.19). TKS assays preformed

in the absence of DABB protein produced an HTAL:PDAL:olivetol ratio of approximately 1:10:2, indicating a proportional two-fold increase in olivetol when compared to TKS assays conduced in the presence of DABB protein. This observation indicates that 1) HTAL production is static in all TKS assays performed, and that 2) OA and olivetol are synthesized using the same TKS-produced intermediate. The discovery of the DABB protein's cyclase activity prompted the naming of this protein to olivetolic acid cyclase (OAC), the enzyme that is the focus of the next chapter.

4. THE CHARACTERIZATION OF OAC

4.1 Introduction

Many polyketide cyclases have been reported in actinomycetes species, and are required during the biosynthesis of antibiotics and pigments (Ames et al., 2011; Shen and Hutchinson, 1993b; Sultana et al., 2004; Thompson et al., 2004). These cyclases are involved in the modification of products derived from type II polyketide synthases, which do not occur in plants, although structurally similar proteins possessing either the Betv1 or DABB motifs have been identified (Bingman et al., 2004; Radauer et al., 2008). These motifs were identified in trichome proteins, and due to their structural relationship with bacterial polyketide cyclases, I initiated an investigation into their putative cyclase activity during the biosynthesis of OA. These experiments led to the discovery of OAC, a DABB protein capable of producing OA when assayed together with TKS (Gagne et al., 2012). The discovery of OAC warranted further investigation of this protein. This chapter describes OAC sequence and phylogenetic analysis, the identification of assay products by LCMS, and experiments designed to characterize OAC. Since type III PKSs are known to exhibit flexibility in substrate specificity, TKS and OAC were also assayed with alternative starter compounds to determine if TKS and OAC could produce alternative resorcylic acids (Wakimoto et al., 2012). Lastly, an addition to the TKS/OAC assays was made by adding recombinant malonyl-CoA synthetase (MCS) derived from *Rhizobium leguminosarum* in hope of augmenting OA biosynthesis (Zhang and Tang, 2009). In summary, the experiments presented in this chapter have helped optimize the TKS/OAC reaction, and provide valuable insight into the biosynthesis of OA by the TKS/OAC reaction.

4.2 Materials and Methods

4.2.1 Phylogenetic Analysis of OAC

The nucleic acid sequence (derived from sequencing) of OAC and its deduced protein sequence were used to conduct eight blast searches against a number of plant protein databases (described in Section 4.3.2). Initially > 150 protein sequences were obtained, but a third of the sequences were discarded since a limit of four representatives from any one species was applied. Representative sequences were arbitrarily chosen. A total of 93 OAC-like sequences (including OAC) were aligned using ClustalW provided in Mega5 (Tamura et al., 2011) with parameters as described in Appendix E. A Maximum Likelihood Phylogenic Tree was constructed using the Mega5 software and alignment file using the settings described in Appendix E. A thousand bootstraps were performed during the analysis.

4.2.2 Biochemical Characterization of OAC

Assays were performed using 100 pmol TKS and 1,000 pmol OAC in 50 µL volumes containing 0.2 mM hexanoyl-CoA, 0.6 mM malonyl-CoA, 50 mM HEPES [pH 7.0], and 5 mM DTT. Reactions were maintained at 30°C for 60 min before being extracted and analyzed by LCMS as described in Section 3.3.1. The following subsections note differences in these methods. All reactions were performed in triplicate.

4.2.2.1 TKS/OAC Ratios

Assays used 10 mM DTT and a range of 250 to 12,500 pmol OAC. Reactions were maintained at 37°C for 60 min. Reactions were analyzed by HPLC as described in Section 3.3.1.

4.2.2.2 Temperature

Assays used 0.5 mM malonyl-CoA. Mixtures were maintained at 10, 15, 20, 25 or 30°C for 5 min before starting the reaction. Reactions were held at the appropriate temperature for 60 min.

4.2.2.3 Buffers

Reactions were conducted in 50 mM buffers including HEPES, TRIS, sodium phosphate, and PIPES. Reactions were incubated for 60 min at 10°C.

4.2.2.4 Concentration of Buffer

The concentration of HEPES used was 12.5, 25, 50, and 100 mM HEPES [pH 7.0].

4.2.2.5 pH

Assays were conducted in: HEPES pH [6.8 to 8.2]; phosphate pH [6.0 to 7.5] which used 0.5 mM malonyl-CoA and were conducted at 10°C; ADA pH [5.7 to 7.8] as described in Section 3.1.3.4 except for the addition of 1000 pmol OAC; and citrate pH [5.0 to 6.0] which used 20 mM buffer. Assays conducted in citrate were run in parallel to reaction performed in 20 mM HEPES pH [7.0] and 20 mM MOPSO pH [7.0]. Citrate buffers were prepared according to (Gomori, 1955).

4.2.2.6 Metals

Reactions were performed in 20 mM HEPES [pH 7.0], and 5 mM of $MgCl_2$, $MnCl_2$, $ZnCl_2$, $FeCl_2$, $CoCl_2$, $CaCl_2$, $CuCl_2$, or water. All metal solutions were freshly prepared from 50 mM stock solutions. Analysis was performed by HPLC as described in Section 3.3.1.

4.2.2.7 EDTA

Reactions were performed in 20 mM HEPES [pH 7.0] in the presence or absence of 10 mM EDTA. Reactions used 200 pmol TKS and 200 pmol OAC, and were maintained at 20°C for 1 h.

4.2.2.8 Malonyl-CoA Synthetase

A 1 mL aliquot of Talon-purified recombinant malonyl-CoA synthetase (MCS) (10 mg/mL) was provided by Mr. Enwu Liu (Zhang and Tang, 2009). Assays using MCS were conducted in 50 µL containing 20 mM HEPES [pH 7.0], 0.2 mM CoA, 0.4 mM ATP, 2.5 mM $MgCl_2$, 8 mM sodium malonate, 5 mM DTT, and 0.2 mM hexanoyl-CoA. Assays were started by adding the appropriate mixture of 5 µg TKS, 10 µg MCS, and 5 µg OAC. Reactions that did not use MCS were conducted using 20 mM HEPES [pH 7.0] as described in Section 4.2.2. Assays were performed at 20°C and were analyzed by LCMS as described in Section 3.3.1.

4.2.2.9 Alternative Substrate Assays

4.2.2.9.1 Free CoA

Assays used 0, 0.25, 0.50, 0.75, or 1.0 mM excess CoA esters. Reactions were started by adding a mixture of 200 pmol TKS and 1,000 pmol OAC and were incubated at 37°C for 1 h.

4.2.2.9.2 TKS products

Assays were performed using 1.0 nmol of OAC in 50 mM HEPES [pH 7.0] and either 1 mM HTAL, 1 mM PDAL, 1 mM olivetol, or 1 mM HTAL + 1 mM CoA and proceeded for 1 h at RT.

4.2.2.9.3 Butyryl-CoA, Octanoyl-CoA, and Acetyl-CoA

Assays used 0.2 mM butyryl-CoA, octanoyl-CoA, or acetyl-CoA instead of hexanoyl-CoA as a starter molecule. Reactions also used 200 pmol TKS, ±1,000 pmol OAC, 0.6 mM malonyl-CoA, and 20 mM HEPES [pH 7.0] and were incubated for 30 min at 30°C. Reactions were prepared

for LCMS analysis as described in Section 3.3.1. Assays using butyryl-CoA were separated over a 30 min gradient run that started with (v/v) 90% water, 10% acetonitrile, and 0.05% formic acid and ended with (v/v) 80% water, 20% acetonitrile and 0.05% formic acid. Assays using octanoyl-CoA were separated over a 25 min run that started with (v/v) 98% water, 2% acetonitrile, and 0.05% formic acid and ended with (v/v) 2% water, 98% acetonitrile and 0.05% formic acid. Assays using acetyl-CoA were analyzed by LCMS as described in Section 3.3.1. All flow rates were maintained at 0.2 mL/min with a 30°C column temperature. Products were monitored at 270 nm and were analyzed using the Empower2 software suite. SIR was conducted as described in Section 3.1.3; SIR m/z values are described in Section 4.3.10.

4.3 Results

4.3.1 Olivetolic Acid Cyclase

OAC is a 101 amino acid protein with a predicted molecular weight of 12 kDa. OAC was initially identified in the CL2contig1 ORF of the CSATR1JP unigenes library previously described by Stout et al. (2012), and cloned into the pET-100 vector resulting in a new construct with a calculated molecular weight of 18.7 kDa. At the amino acid level, OAC shares 48% identity with a heat stable protein from *Arabidopsis thaliana* named AtHS1 (annotated NP_566569), and 38% identity to *Populus tremula Pt*SP1 (annotated CAC34953), which is upregulated during stress responses (Park et al., 2007). The crystal structure for AtHS1 has been solved and confirms the presence of dimeric α + β barrel domains (Bingman et al., 2004). Although AtHS1 was shown to possess antifungal and antibacterial activity, no specific physiological function has been attributed to this protein (Park et al., 2007). Similarly, it is not known why the heat-stable SP1 is upregulated during heat, salt and stress responses, yet *A. thaliana* plants expressing transgenic SP1 have been shown to be heat-tolerant (Zhu et al., 2008). Structurally, the OAC protein sequence was used to query the Interpro web resource, revealing a plethora of DABB proteins with many different functions. Of specific relevance is the TcmI protein from *Streptomyces glaucescens* (annotated AAA67513), which assists in the cyclization of TcmF2 to TcmF1 during the biosynthesis of tetracenomycin C, an antitumor/antibiotic aromatic polyketide (Thompson et al., 2004). The TcmI DABB domain belongs to a superfamily of actinobacterial proteins known as the PKCs (Interpro IPR006765; CDD cl04663). It was partially this structural relationship that initially implicated OAC as a candidate polyketide cyclase. Structurally, the OAC DABB domain belongs to a different family of proteins than

TcmI, called the stress-responsive alpha beta barrel (SRABB) proteins by Interpro (IPR013097), or simply as DABB proteins by the CDD (cl15807). Conversely, DABB proteins are recognized by Interpro as a superfamily of structurally similar proteins (IPR011008) which include SRABBs and PKCs, as well as many other structurally similar protein families.

4.3.1.1 Expression and Purification of OAC

Originally, OAC was cloned in the pET100 vector, expressed in *E. coli* BL21, and purified using IMAC. Although this produced active protein to be processed, low amounts of impure protein were obtained. The low expression of OAC in the pET100 vector led to the construction of the pHIS8-OAC plasmid, but this change did not substantially increase protein yields, which were only between 1-2 mg/L culture. Furthermore, the expression and purification of OAC from the pHIS8-OAC vector resulted in protein fragmentation attributed to OAC degradation. During SDS-PAGE, IMAC-purified OAC migrates as a (ca.) 19 kDa band next to a less intense 17 kDa band (Figure 4.1). Both bands were analyzed by MS/MS, with results indicating that both bands correspond to OAC, suggesting that the 17 kDa contaminant was degraded OAC. Various attempts to further purify OAC using chromatographic techniques including size exclusion, ion exchange and hydrophobic interaction chromatography had little success in removing the degradation product. Assays that were performed throughout this chapter used OAC that was purified using minimal amounts of Talon and, with the exception of the degradation product, resulted in a purified active product (Figure 4.1).

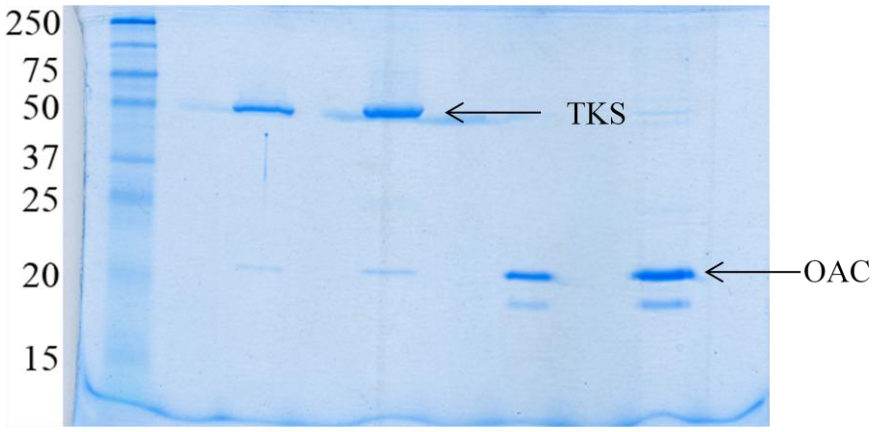


Figure 4.1 SDS-PAGE of Talon purified TKS and OAC.

A 15% polyacrylamide gel was used to resolve 1.25 µg TKS (lane 3), 3.75 µg of TKS (lane 5), 1.25 µg OAC (lane 7) and 3.75 µg OAC (lane 9) next to 10 µL Prestained Protein ladder (lane 1) at 200V for 45 min. OAC did show a contaminating band at (ca.) 17 kDa as seen in lanes 7 and 9. Numbers on the left of the gel are the corresponding protein masses (kDa).

4.3.2 The Evolutionary Relationship of OAC

A search for OAC homologs was conducted using the OAC protein sequence during a variety of blast queries in a number of plant sequence databases, namely the NCBI Transcriptome Shotgun Assembly Sequence Database (Benson et al., 2000), the Cannabis Genome database (Van Bakel et al., 2011), TIGR Plant Transcript Assemblies (Childs et al., 2007), the *Solanum lycopersicum* project (Tomato Genome Consortium, 2012), GreenPhylDB (Conte et al., 2008), and Phytozome (Goodstein et al., 2012). A total of 93 protein sequences with an expected threshold $<1x10^{-10}$ were aligned using MEGA5 software's ClustalW implementation, and then used to construct a Maximum Likelihood phylogenetic tree (Tamura et al., 2011) (Figure 4.2). Names of sequences are included in Appendix F. A limit of four homologs from any given species to be included in the alignment was applied when too many sequences from one species were detected. Sequences used in the alignment were obtained from BLAST searches against higher plant genomes, but

also included *Cyanidioschyzon merolae*, *Chlamydomonas reinhardtii*, and *Ostreococcus tauri* genomes, although no OAC homolog within an expected threshold of $1x10^{-2}$ could be detected within these algal genomes. Phylogenic analysis indicates that OAC homologs form two distinct clades that further subdivide into monocot and dicot lineages (Figure 4.3 and 4.4). Broader searches for OAC homologs against GenBank non-redundant sequences clearly detected their existence in a variety of bacteria including Proteobacteria, Firmicutes, Enterobacteriaceae, and Sphingobacteria with E-values less than $1x10^{-14}$.

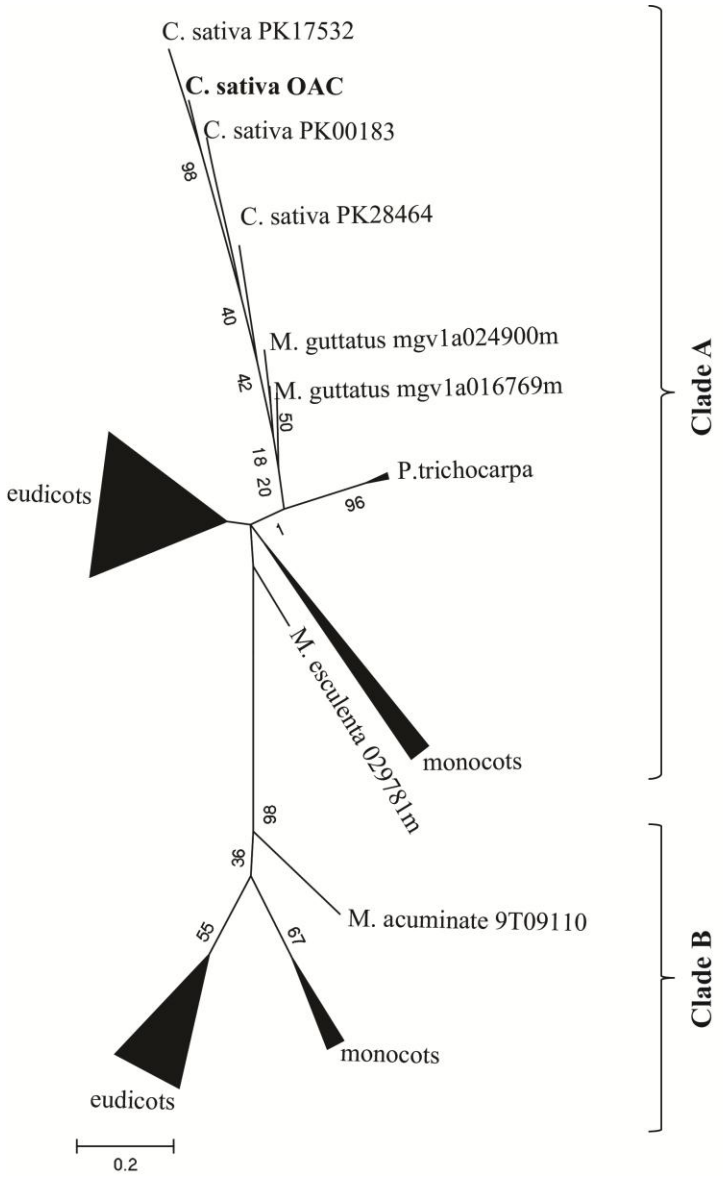


Figure 4.2 Phylogenetic relationship of plant OAC homologs.

The evolutionary history was inferred using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992). The bootstrap consensus tree inferred from 1,000 replicates represents the phylogenetic relationship of OAC (Felsenstein, 1985). Bootstrap values are shown along branch lengths. The analysis involved 93 amino acid sequences (Appendix F). There were a total of 58 positions in the final dataset. Analyses were conducted in MEGA5 (Tamura et al., 2011). Two distinct clades (A and B) divide the homologs, which can be further divided into eudicot and monocot groups. The scale bar indicates the genetic distance of 0.2 substitution/site.

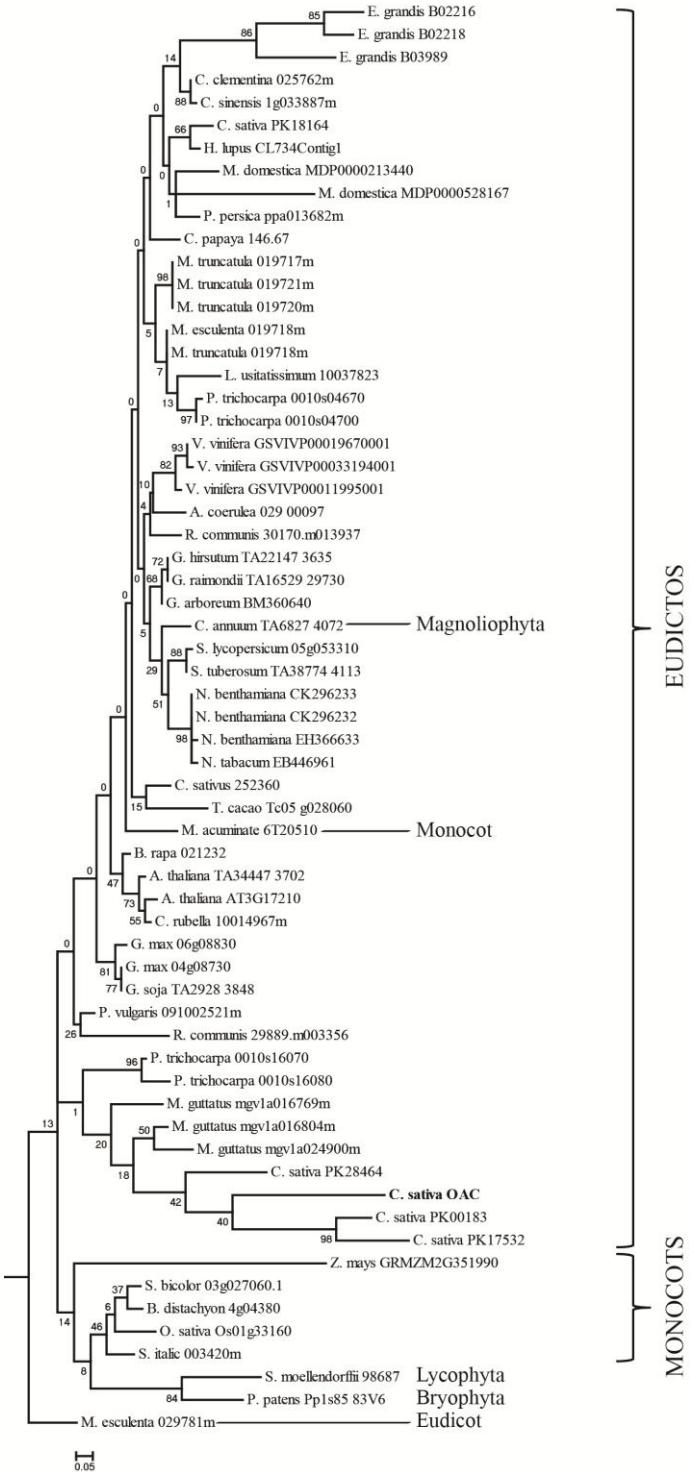


Figure 4.3 Phylogenetic relationship of OAC plant homologs in Clade A.

Plant OAC homologs found in clade A are shown. The scale bar indicates the genetic distance of 0.05 substitution/site.

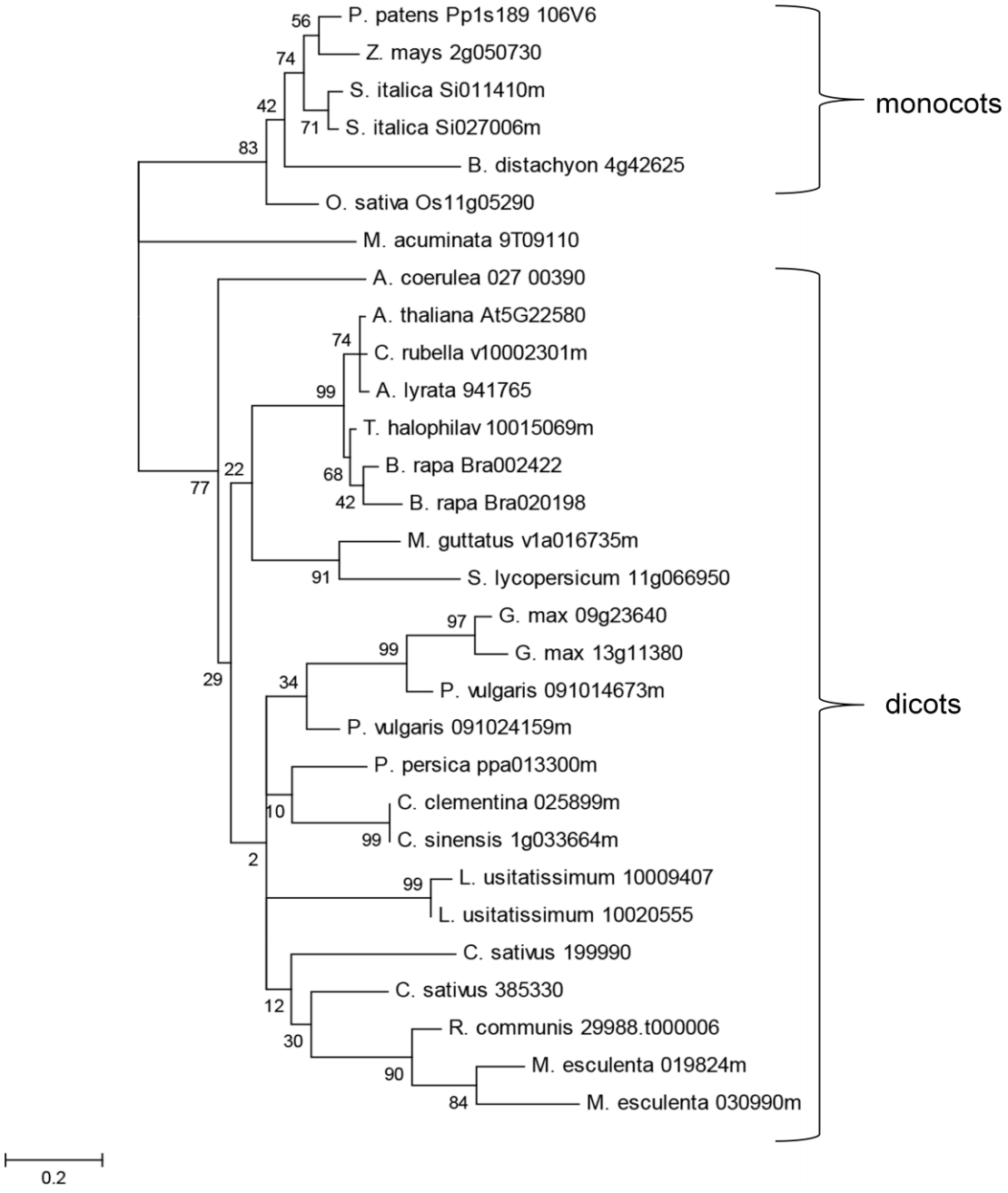


Figure 4.4 Phylogenetic relationship of OAC plant homologs in Clade B.

OAC homologs that form the 'B' clade division are shown. These homologs can be further subdivided into monocot and eudicot classifications. The scale bar indicates the genetic distance of 0.2 substitution/site.

4.3.3 The TKS/OAC Reaction Produces Olivetolic Acid

TKS assays conducted in the presence of OAC produced four compounds identifiable by comparing their retention times, UV spectra, and masses to authentic standards during LCMS analysis (Figure 4.5). The four products included HTAL, PDAL, olivetol, and OA, which were also confirmed by single ion monitoring in electrospray positive ion mode at 224.95 m/z for HTAL and OA, 182.96 m/z for PDAL, and 180.91 m/z for olivetol (Figure 4.6). Product ratios (*i.e.* chemoselectivity) were affected by assay conditions including temperature, pH, and substrate concentrations and are discussed in the appropriate Sections. Product ratios generally mirrored product accumulation of TKS assays performed in the absence of OAC, with the exception of olivetol which accumulated to a lesser degree in TKS assays performed in the presence of OAC, and of course OA which does not accumulate when TKS is assayed alone (Figure 4.7). Typically, product ratios of HTAL:PDAL:OA:olivetol were in *circa* 1:10:1:1 throughout this work, but these values did fluctuate dramatically. Once reaction conditions were optimized for OA production, a product ratio of 1:7:2:1 was obtained.

The LCMS chromatograms of standards and the TKS/OAC assay presented in Figures 4.5 and 4.6 can be directly compared to the chromatograms presented in Chapter 3 (Figure 3.4 and 3.5); they were obtained from experiments conducted in parallel. Note that the chromatograms shown are derived from reactions conducted once optimized reaction conditions were determined.

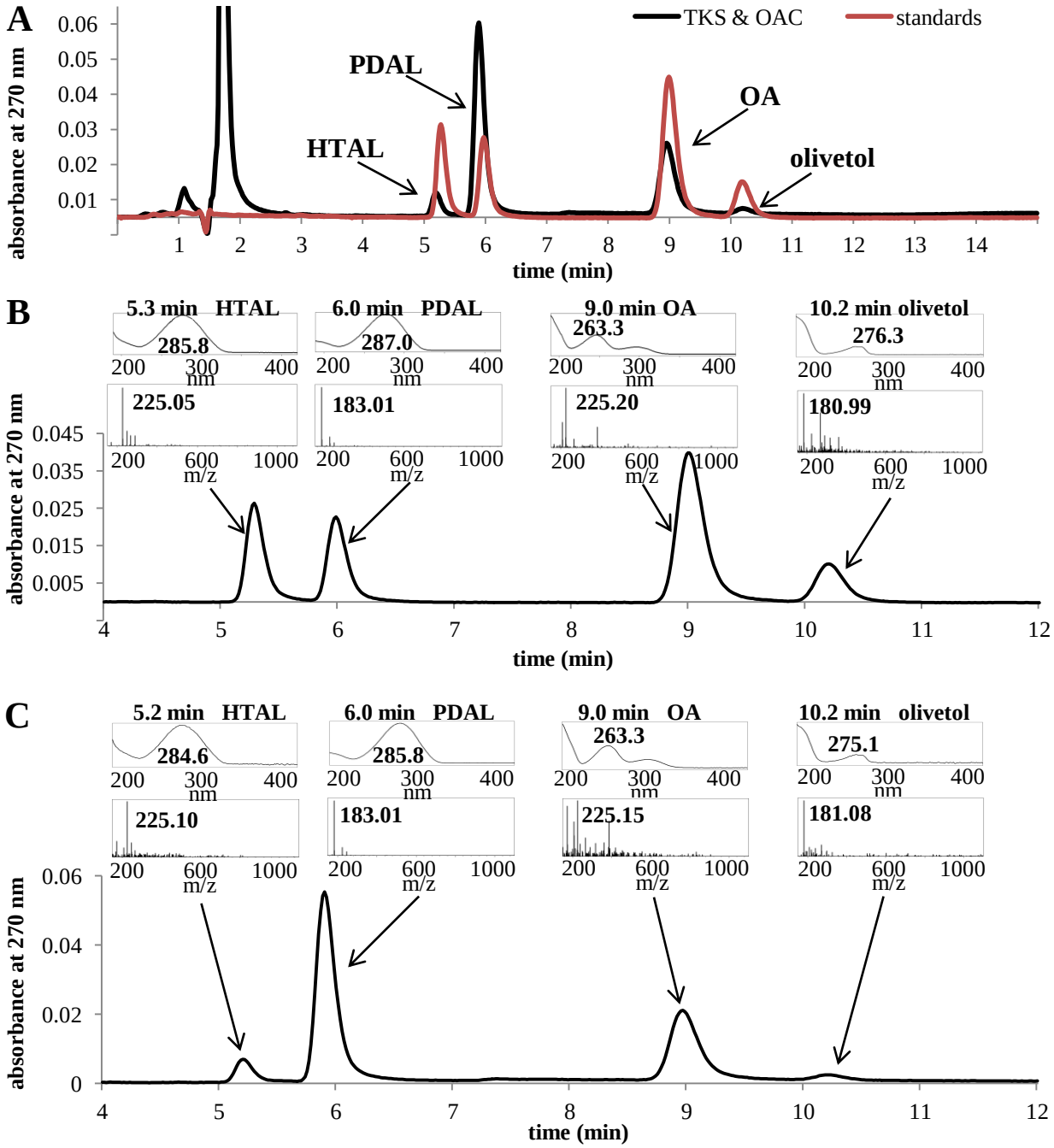


Figure 4.5 LCMS of standards and TKS/OAC assay products.

(A) Overlayed LC chromatograms for TKS/OAC assay products (black) and standards (red) indicate the similarity in product retention times. (B) An LCMS chromatogram of the standards used in this study indicate the retention times, UV spectra and extracted ion profiles for HTAL (5.2 min; UV_{max} 285 nm; m/z 225), PDAL (5.6 min; UV_{max} 287 nm; m/z 183), OA (9.0 min; UV_{max} 263 nm; m/z 225.), and olivetol (10.2 min; UV_{max} 276 nm; m/z 181) and were used to confirm the identities of (C) TKS/OAC assay products shown in a similar LCMS trace.

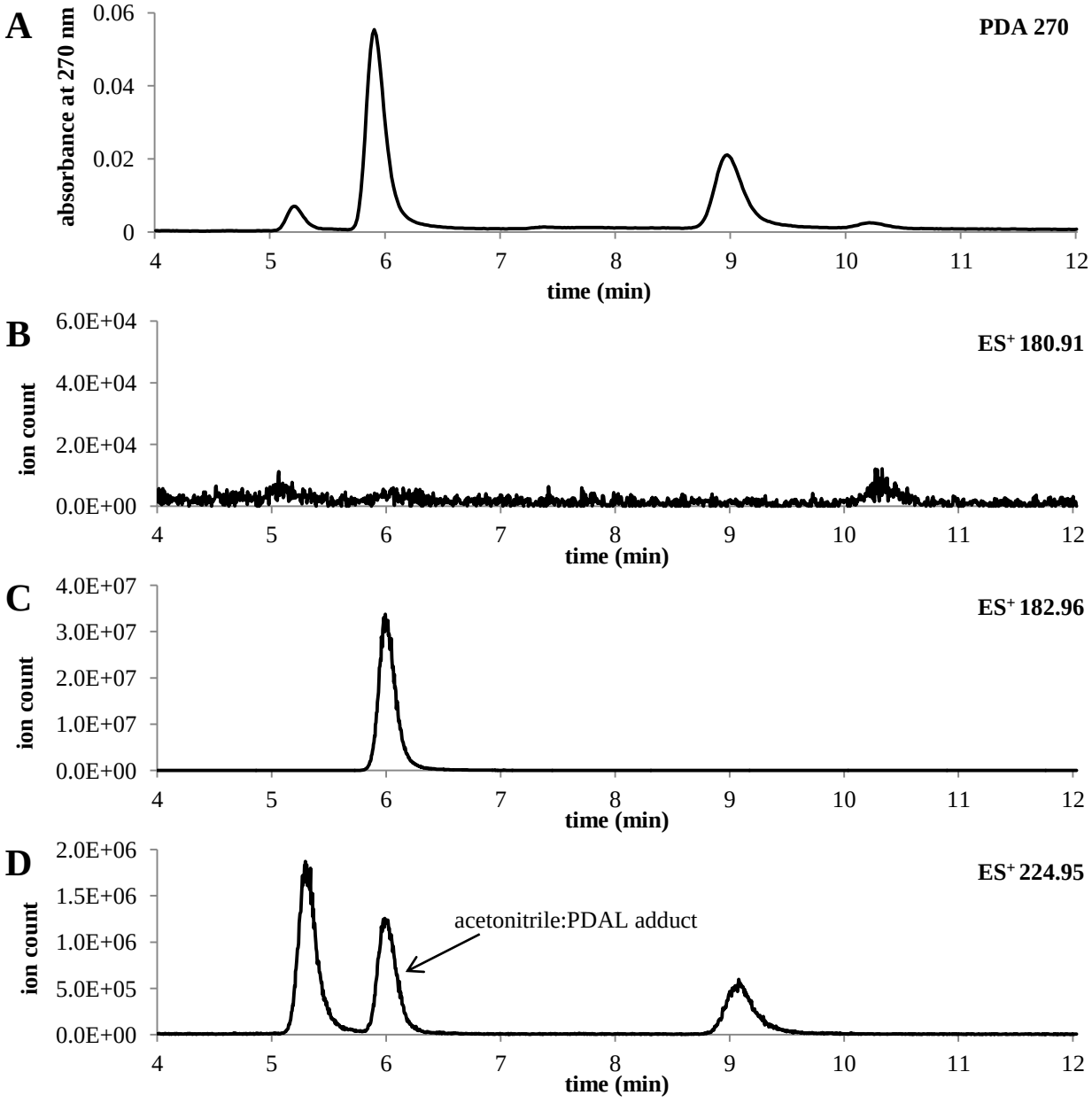


Figure 4.6 Selected ion monitoring of TKS/OAC assay products.

LCMS was used to confirm the identity of HTAL, PDAL and olivetol produced during TKS assays. (A) PDA results align well with selected ion monitoring in ES^+ mode which was capable of detecting (B) olivetol (10.2 min; m/z 180.91), (C) PDAL (6.0 min; m/z 182.96), (D) HTAL (5.2 min; m/z 224.95) as well as olivetolic acid (9.0 min; m/z 224.95). Another peak (6.0 min) detected during (D) selected ion monitoring at 224.95 m/z is attributed to acetonitrile:PDAL adduct formation.

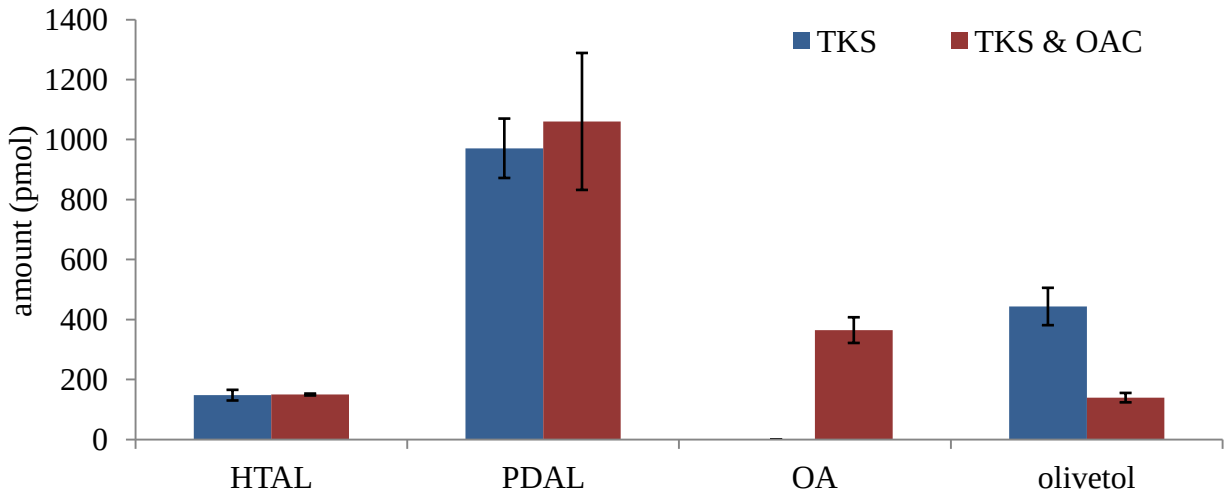


Figure 4.7 A comparison of TKS and TKS/OAC assay products.

TKS assayed in the presence of OAC resulted in the accumulation of OA apparently at the cost of olivetol accumulation. Error bars are the SD where n=3.

4.3.4 Optimum Ratio of TKS and OAC

Experiments were conducted to establish the ratio of TKS to OAC required to optimize the production of OA. Additionally, this work tested if a stoichiometric relationship exists between TKS and OAC, which could suggest if a physical interaction occurs between these enzymes. The optimum amount of starting enzyme was determined by assaying 200 pmol (ca. 10 µg) TKS in the presence of 250 to 12,500 pmol (ca. 230 µg) of OAC (Figure 4.8). The largest amount of OA accumulation (538 pmol) during these assays was observed when 200 pmol TKS was assayed in the presence of 7,500 pmol (ca. 140 µg) OAC. OAC activity appears to level off when amounts greater than 7,500 pmol of OAC was assayed. The accumulation of OA is inversely proportional to the accumulation of olivetol, and has a correlation coefficient of -0.978 (P value of 0.00001).

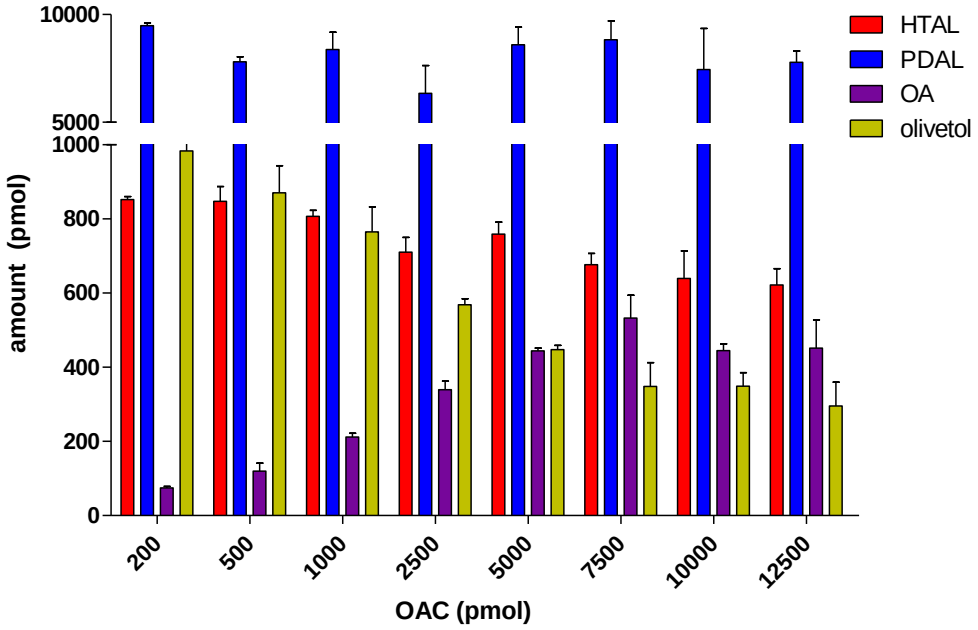


Figure 4.8 TKS assays conducted with variable amounts of OAC.

Product accumulation during TKS/OAC assays was affected by the concentration of OAC. Constant amounts of 200 pmol TKS were assayed in the presence of 250 to 12,500 pmol OAC. Error bars are the SD where n=3. Vertical axis is split to show smaller amounts of products.

4.3.5 OAC Activity as a Function of Temperature

These experiments were conducted to examine the effect temperature has on the TKS/OAC reaction and to determine the optimum temperature at which OA production occurs. Temperature influenced both the product ratios and total amounts of products from TKS/OAC assays. Results indicate that OA production increases at lower (10 to 20°C) temperatures, while HTAL, PDAL and olivetol production decreased (Figure 4.9). Although previous TKS assays indicated an increase in product formation at (ca.) 40°C (Figure 3.9), the largest amount of OA detected in TKS/OAC assays occurred at 15°C and was not significant at 10°C (z for 95% CI= 1.96).

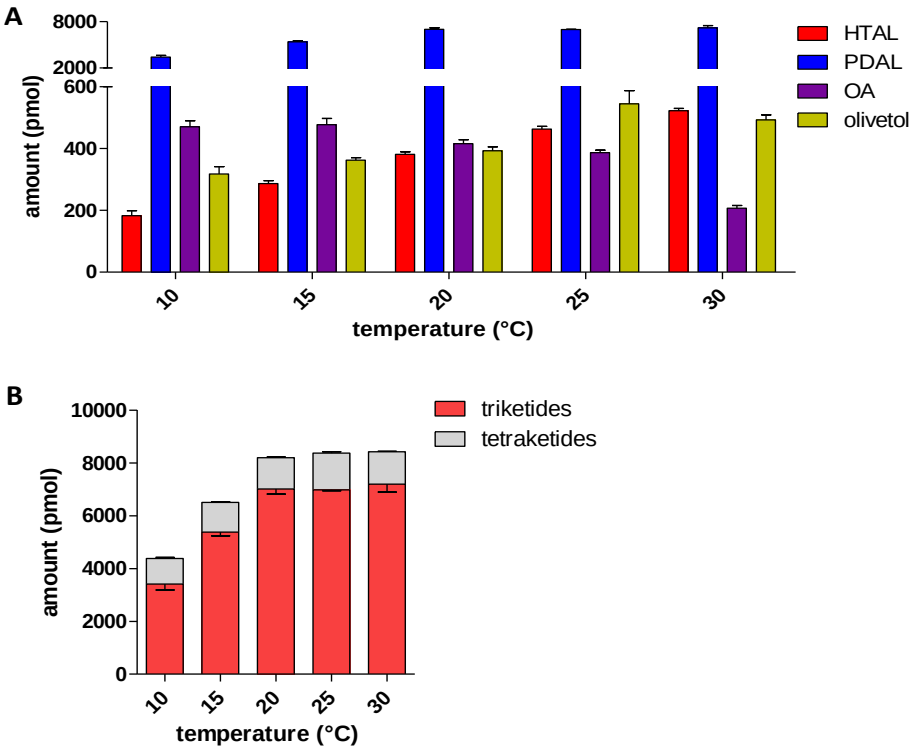


Figure 4.9 The activity of TKS/OAC as a function of temperature.

(A) As temperature is increased, so is the production of HTAL, PDAL, and olivetol. Conversely, OAC activity increases at lower (10 to 20°C) temperatures. Vertical axis is split to show smaller amounts of products. (B) The total amount of products is increased at temperatures of 20 to 30°C compared to lower temperatures. Error bars are the SD where n=3.

4.3.6 OAC Activity in Different Buffer Systems

The activity of both TKS and OAC was strongly influenced by the buffer in which assays were conducted. Assays were performed in triplicate using (final) concentrations of 50 mM HEPES, TRIS, phosphate, or PIPES at pH 7.0, and product accumulation was compared (Figure 4.10). HEPES appears a better buffer than TRIS, phosphate, or PIPES, and was capable of reducing the amount of PDAL production while increasing overall tetraketide (HTAL, OA, and olivetol) production. Assays conducted in the presence of HEPES buffer compared to other buffering systems accumulated more than two times the amount of OA (1,100 pmol), and more than three times the amounts of HTAL (569 pmol) and olivetol (1,061 pmol).

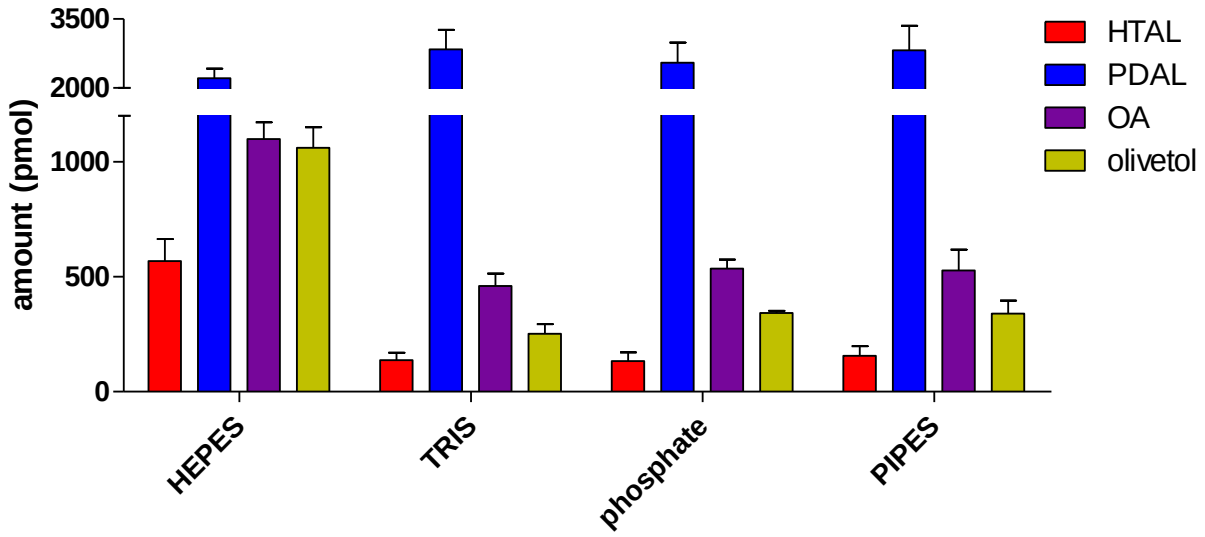


Figure 4.10 TKS/OAC assays conducted in different buffers.
When TKS and OAC were assayed in HEPES, an increase in tetraketide (HTAL, olivetol, OA) production was observed compared to assays conducted in TRIS, phosphate or PIPES buffer. All assays were conducted in the presence of 50 mM buffer [pH 7.0]. Vertical axis is split to show smaller amounts of products. Error bars are the SD where n=3.

4.3.6.1 OAC Activity in Varying Concentration of HEPES pH 7.0

Product accumulation from TKS/OAC assays was also investigated using different molarities of HEPES buffer, where final concentration of HEPES at pH 7.0 were 12.5, 25, 50, and 100 mM (Figure 4.11A). This experiment was conducted to determine if HEPES concentration affected OA production. The largest amounts of tetraketides were detected in assays that used 12.5 mM HEPES including HTAL (573 pmol), OA (621 pmol), and olivetol (574 pmol). Total polyketide production was unaffected by the HEPES concentrations used (Figure 4.11B)

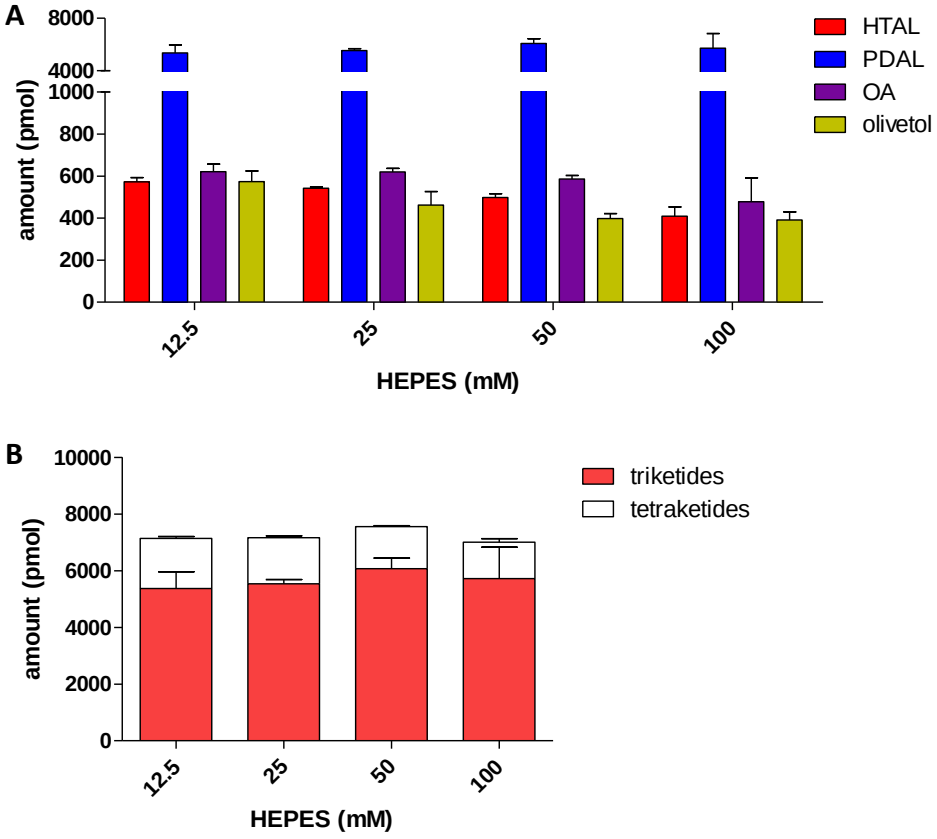


Figure 4.11 Activity of TKS/OAC in variable concentrations of HEPES.

(A) A comparison of product accumulation during TKS and OAC assays conducted in the presence of 12.5 – 100 mM HEPES buffer at pH 7.0. (B) The total amount of PDAL (triketide) is compared to the sum of tetraketides (HTAL, OA, and olivetol). Vertical axis is split to show smaller amounts of products. Error bars are the SD where n=3.

4.3.7 OAC Activity as a Function of pH

The accumulation of TKS/OAC assay products was monitored over a pH range in a variety of buffers including HEPES, phosphate, ADA, and citrate. Given that a marked increase in tetraketide production was observed in assays conducted in HEPES buffer, product accumulation at different pHs was analyzed using this buffer. Unfortunately, the effective pH range of HEPES is too narrow (6.8 to 8.2) to accomodate for more acidic environments. Monitoring of reaction products at pH lower than 6.8 was later achieved using ADA and phosphate as buffering agents. Taura et al. (2009) used citrate [pH 5.5] as a buffer for OLS, and this led to the further analysis of TKS/OAC products in this buffer, which was directly compared to TKS/OAC assays conducted in HEPES and MOPSO buffers.

4.3.7.1 OAC Activity in HEPES pH 6.8 to 8.2

Assays conducted in HEPES buffers within its full pH range indicate that maximal production of OA, as well as HTAL and olivetol occurs at a lower pH than HEPES could accommodate (Figure 4.12). The largest amount of OA (704 pmol) was detected in assays conducted at pH 6.8.

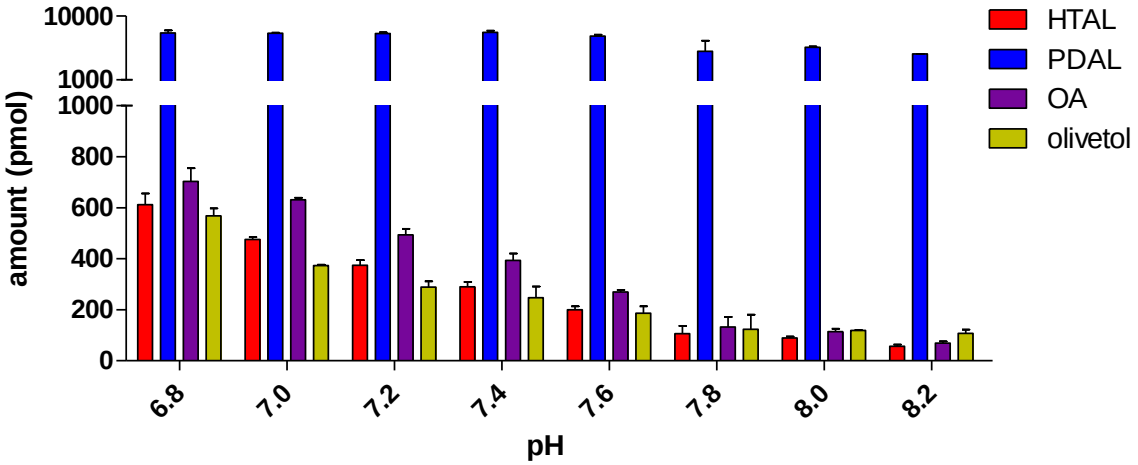


Figure 4.12 Activity of TKS/OAC in HEPES buffer pH 6.8 to 8.2.
A pH profile using 50 mM HEPES was developed by assaying 100 pmol TKS and 1,000 pmol OAC over the full buffering range of HEPES (6.8 to 8.2). Vertical axis is split to show smaller amounts of products. Error bars are the SD where n=3.

4.3.7.2 OAC Activity in Phosphate Buffer pH 6.0 to 8.0

Although it was previously shown that phosphate decreases TKS and OAC activity (Figure 4.10), the buffering range of phosphate allowed for a broader examination of pH and its effect on activity (Figure 4.13). These assays suggest an optimum pH 6.5 for the production of OA (442 pmol) and HTAL (173 pmol), pH 7.5 for the production of PDAL (2650 pmol), and pH 6.0 for olivetol (704 pmol). The lower pH at which olivetol production is increased suggests that olivetol production could be augmented at an even lower pH.

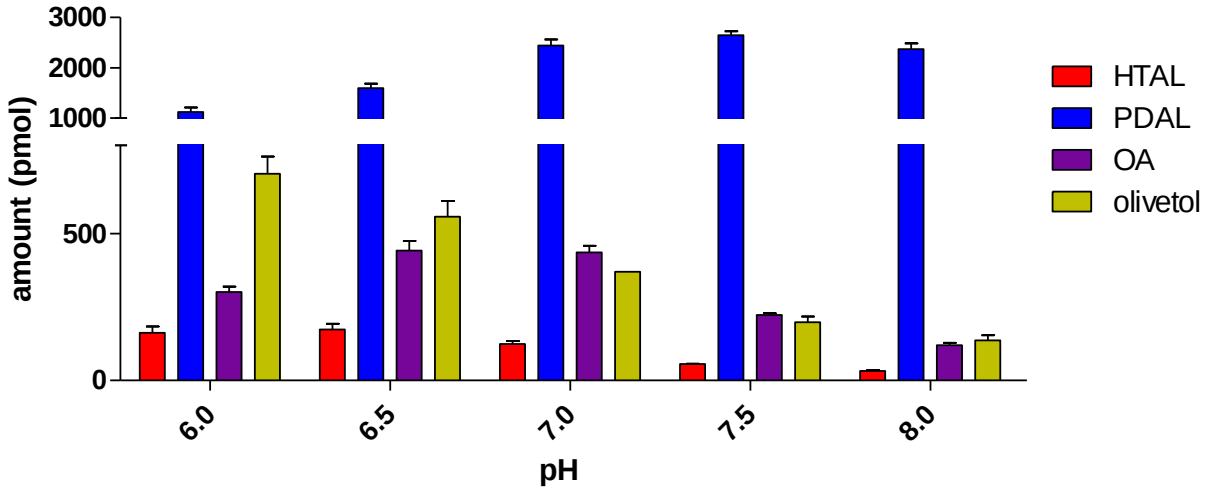


Figure 4.13 Activity of TKS/OAC in phosphate buffer pH 6.0 to 8.0.
A pH profile using 50 mM phosphate was obtained by assaying 100 pmol TKS and 1,000 pmol OAC in the presence of malonyl-CoA and hexanoyl-CoA over a broad pH range (6.0 to 8.0). Vertical axis is split to show smaller amounts of products. Error bars are the SD where n=3.

4.3.7.3 in ADA Buffer pH 5.7 to 7.8

The activity of TKS and OAC were also examined using ADA buffer at pH intervals of 0.2 between the range of 5.7 to 7.8 (Figure 4.14). This pH profile is the most comprehensive of all pH curves obtained for TKS/OAC assays and indicates a similar trend as the other pH profiles. Maximal HTAL (153 pmol) and OA (550 pmol) production occurred at pH 6.4, whereas olivetol (83 pmol) accumulation was highest at pH 6.0, and that of PDAL (pmol 1,361) was at pH 7.2. These ADA pH profiles were obtained during experiments conducted in parallel with those presented in Section 3.2.2.1, where TKS was assayed alone.

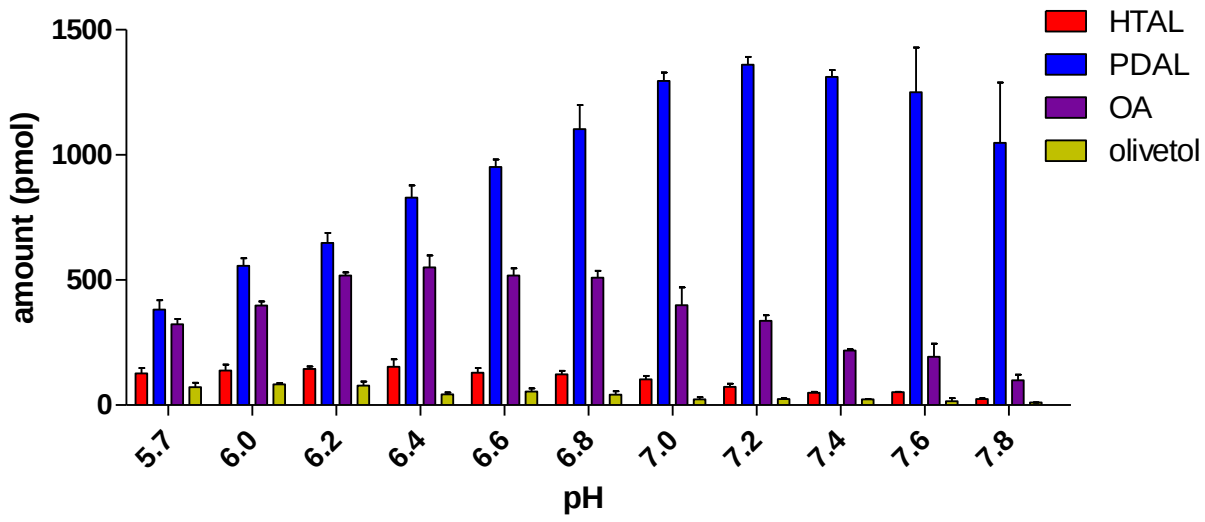


Figure 4.14 Activity of TKS/OAC in ADA buffer pH 5.7 to 7.8.

A pH profile using 50 mM ADA was obtained by assaying 100 pmol TKS and 1,000 pmol OAC in the presence of malonyl-CoA and hexanoyl-CoA over a broad pH range (5.7 to 7.8). Vertical axis is split to show smaller amounts of products. Error bars are the SD where n=3.

4.3.7.4 in Citrate Buffer pH 5.0 to 6.0

TKS and OAC were assayed in the presence of 20 mM citrate at pH 5.0, 5.25, 5.5, 5.75, and 6.0 and product accumulation was directly compared to assays conducted in 20 mM HEPES at pH 7.0 and in 20 mM MOPSO at pH 7.0 (Figure 4.15). The largest amount of OA was detected from assays performed in MOPSO at pH 7.0, where OA accumulated to 754 pmol. The second largest OA accumulation was from assays conducted in HEPES at pH 7.0 where OA accumulated to 605 pmol. The largest amounts of olivetol (1,508 pmol) and HTAL (523 pmol) were detected in assays conducted in citrate buffer at pH 5.75, although assays conducted in citrate generated the least amount of OA (456 pmol), which was maximally produced in citrate at pH 6.0.

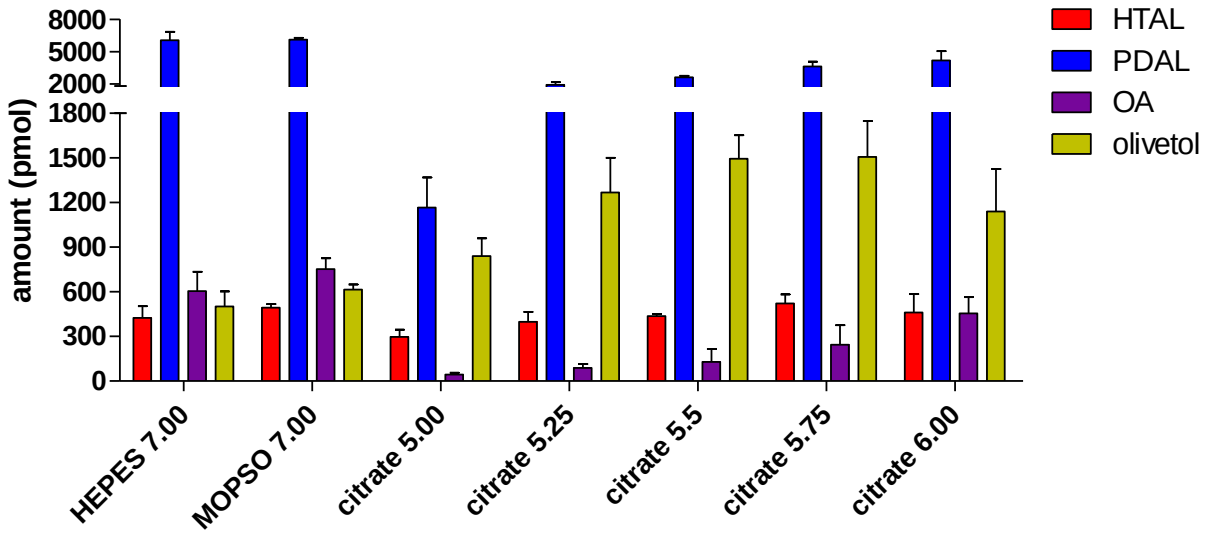


Figure 4.15 Activity of TKS/OAC in citrate buffer pH 5.0 to 6.0.

A pH curve using 20 mM citrate was obtained by assaying 100 pmol TKS and 1,000 pmol OAC in a pH range of 5.0 to 6.0 and is directly compared to assays conducted in HEPES [pH 7.0] and MOPSO [pH 7.0]. Vertical axis is split to show smaller amounts of products. Error bars are the SD where n=3.

4.3.8 OAC Activity in the Presence of Metals

These experiments were conducted to determine if metals had an effect on TKS and OAC activity. The effect that metals have on OAC activity was determined by assaying the enzymes in the presence of Mg^{2+}, Ca^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Cu^{2+} and Zn^{2+} (Figure 4.16). Whereas alkaline earth metals had a relatively small effect on TKS/OAC activity, the transition metals strongly affected the product profile of the TKS/OAC reaction. Notably, Fe^{2+}, Co^{2+}, Zn^{2+} and Cu^{2+} reduced the amount of HTAL, PDAL, and OA to an average of 38%, 20%, and 23%, respectively when compared to TKS/OAC reactions without metals. These same metals also increased the amount of olivetol by an average of 84% when compared to the amount of olivetol produced in TKS/OAC reactions without metals. The largest amount of OA was observed in the absence of metals.

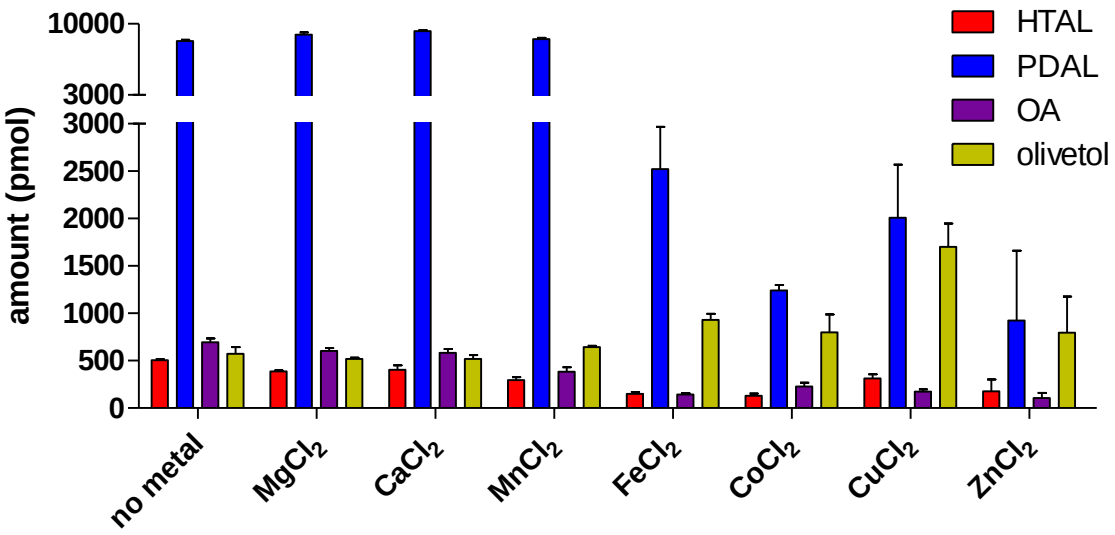


Figure 4.16 The effect of metal on the TKS/OAC reaction.

This figure compares product accumulation in TKS/OAC reactions conducted in the presence and absence of selected metals. Vertical axis is split to show smaller amounts of products. Error bars are the SD where n=3.

4.3.8.1 TKS/OAC Activity in the Presence of EDTA

In addition to surveying OAC activity in the presence of metals, TKS and TKS/OAC were also assayed in the presence and absence of 10 mM EDTA, a hexadentate chelating agent capable of sequestering metal ions (Figure 4.17). These experiments were conducted to examine TKS/OAC activity in the absence of metals. The addition of EDTA to TKS and TKS/OAC assays increased the production of PDAL by about 30%, whereas the production of HTAL and olivetol dropped by 40% and 25%, respectively, when compared to assays performed in the absence of EDTA. The production of OA also diminished in TKS/OAC assays conducted in the presence of EDTA, but only by 8% when compared to similar assays performed in the absence of EDTA.

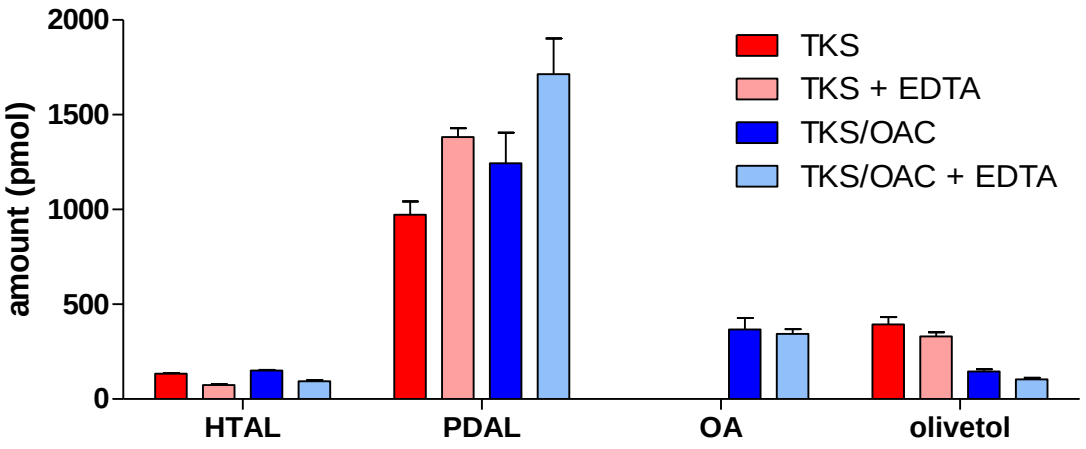


Figure 4.17 EDTA affects the synthesis of TKS/OAC products.

A comparison of reaction products that accumulated during TKS and TKS/OAC assays in the presence and absence of EDTA show a reduction in tetraketide production and an increase in PDAL production in EDTA containing reactions. Error bars are the SD where n=3.

4.3.9 TKS/OAC Activity in the Presence of MCS

Malonyl-CoA synthetase (MCS) was used to produce malonyl-CoA *in vitro* alongside TKS/OAC assays. These assays were performed to compare the activity of TKS and TKS/OAC in the presence and absence of MCS. Although the inclusion of MCS in assays allowed for the production of malonyl-CoA, the reaction required the addition of free CoA, ATP, $MgCl_2$ and sodium malonate. Assays which did not include MCS required malonyl-CoA, but not free CoA, ATP, $MgCl_2$ and sodium malonate. In experiments conducted in parallel, TKS assays that used malonyl-CoA produced an average 32% more HTAL, PDAL and olivetol than TKS assays using MCS (Figure 4.18). Similarly, TKS/OAC assays that used malonyl-CoA showed increases of 53% HTAL, 36% PDAL, 9% olivetol, and 40% OA when compared to TKS/OAC assays that employed MCS.

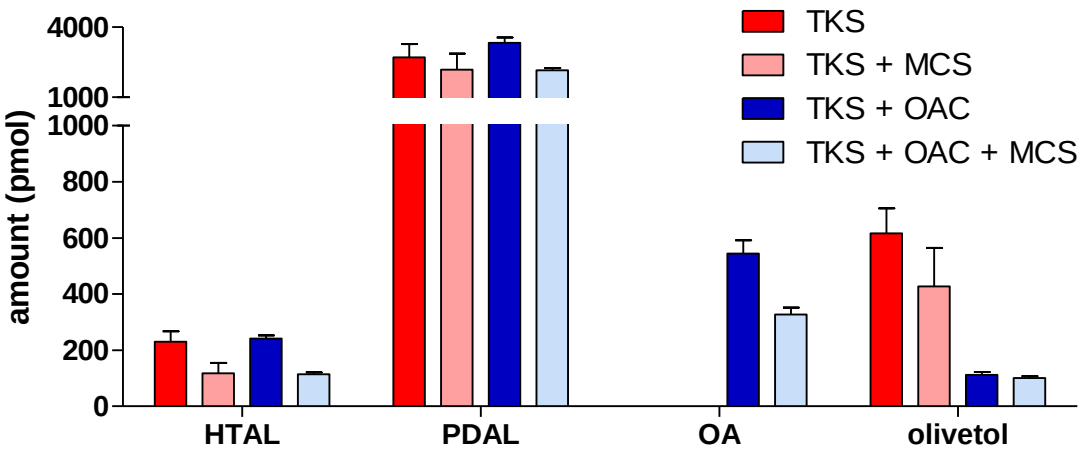


Figure 4.18 TKS/OAC assays conducted in presence and absence of MCS.
TKS and TKS/OAC assays were performed with and without MCS allowing for the comparison of product accumulation. Vertical axis is split to show smaller amounts of products. Error bars are the SD where n=3.

4.3.10 Substrate Utilization in the TKS/OAC Reaction

These experiments were conducted to determine if TKS assay products were suitable substrates for OAC. Free CoA was added to the TKS/OAC reaction to determine if it influenced OA production, and to determine if free CoA could be involved in the acquisition and use of HTAL as a substrate by OAC. OAC was assayed without TKS in the presence of the polyketide products themselves, which included HTAL, PDAL, and olivetol, to determine if OAC was capable of using these compounds as substrates for the production of OA. Additionally, to explore TKS promiscuity, enzyme assays were performed using starter molecules other than hexanoyl-CoA, which included octanoyl-CoA, butyryl-CoA, and acetyl-CoA. TKS assays were conducted with and without OAC using alternative starter compounds and product accumulation was compared.

4.3.10.1 The Effect of Free CoA on TKS/OAC Activity

TKS/OAC assays were conducted in the presence and absence of free CoA and product accumulation was compared (Figure 4.19). When the concentration of free CoA was increased, PDAL production also appeared to increase, but the amount of HTAL, olivetol and OA was diminished. The maximal amount of OA (159 pmol) was produced by TKS/OAC assayed without CoA, whereas the smallest amount of OA (23 pmol) was recovered when TKS/OAC was assayed in the presence of 0.75 mM CoA, resulting in an 85% reduction in OA production.

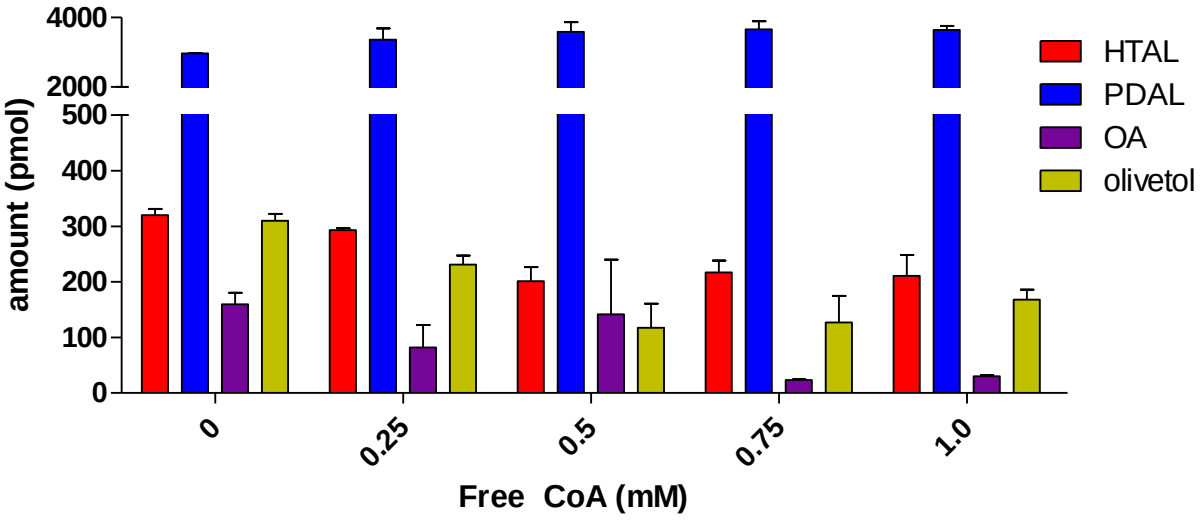


Figure 4.19 TKS/OAC product accumulation in the presence of free CoA.

An increase in PDAL and a decrease in HTAL, OA, and olivetol were detected when TKS/OAC was assayed in the presence of free CoA, when compared to assays conducted in the absence of free CoA. Vertical axis is split to show smaller amounts of products. Error bars are the SD where n=3.

4.3.10.2 TKS Products as Substrates for OAC

The TKS/OAC reaction clearly leads to the production of OA, but the question as to which substrate OAC cyclizes remains unanswered. The unidentified substrate could be a linear tetraketide that spontaneously cyclizes to either HTAL or olivetol in solution, or as a release mechanism from TKS, but studies that have tried to produce reactive triketo acids have failed to report the production of lactones from the cyclization of free acids (Harris and Carney, 1966; Harris and Carney, 1967; Staunton and Weissman, 2001). The existence of linear polyketide intermediates has also been questioned during substrate channeling between CHS and downstream enzymes, such as chalcone reductases (CHRs); whether or not these substrates are enzyme-bound or exist as free intermediates remains unknown (Austin and Noel, 2003). To address the issue of OAC substrate utilization, OAC was assayed in the presence of isolated TKS products (HTAL, PDAL, and olivetol), which could act as substrates *if* OAC possesses ring-opening activity. Additionally, OAC was assayed with 0.5 mM HTAL and either 1) 0.5 mM free CoA, or 2) 200 pmol TKS. The addition of free CoA to these assays was investigated since ring-opening activity could require a cofactor which was present when TKS products are generated *in vitro.* The addition of TKS to assays allowed the investigation as to whether or not TKS could itself be responsible for the potential ring-opening of HTAL, resulting in the production of triketo acid intermediates. When OAC was assayed with either PDAL or olivetol, no other compound could be recovered (traces not shown). OAC assays performed in the presence of 1.0 mM HTAL resulted in the detection of trace amounts of OA during LCMS analysis (Figure 4.20). The addition of free CoA or TKS to OAC assays did not lead to an increase of OA.

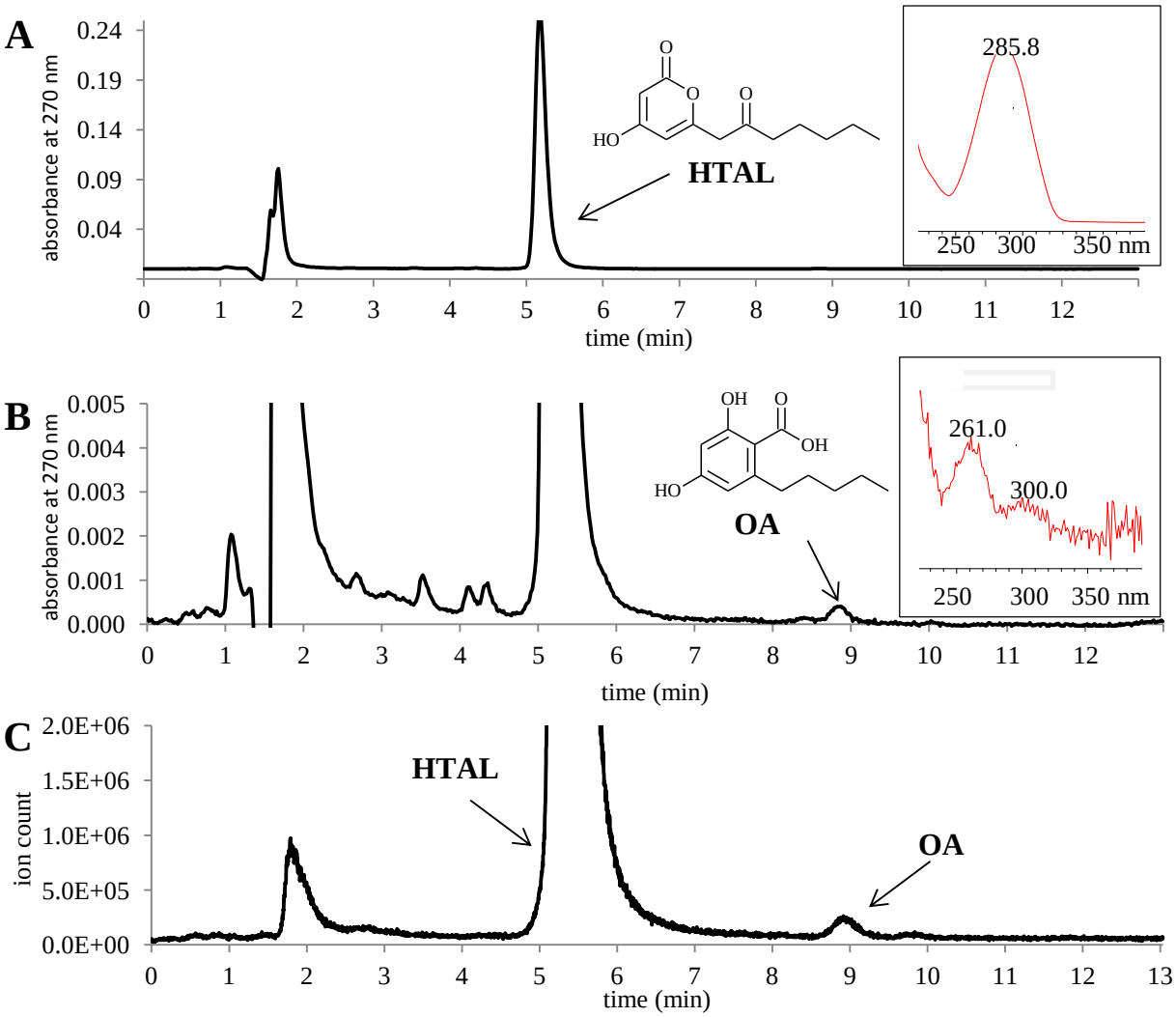


Figure 4.20 LCMS of OAC assays conducted with HTAL.

The LCMS analysis of products generated when OAC was assayed with 0.5 mM HTAL. (A) PDA trace clearly indicates the elution of HTAL at 5.2 min, although no OA is observable at this scale. The UV spectrum for HTAL is shown to the right. (B) Trace amounts of OA are detectable at a UV absorbance scale of 0.005 at 8.9 min. The UV spectrum for OA is shown to the right. (C) Analysis by SIR at m/z of 224.95 in ES^+ mode was capable of detecting HTAL at 5.2 min, as well as trace amounts of OA at 8.9 min.

4.3.10.3 With Butyryl-CoA

TKS was assayed in the presence and absence of OAC using butyryl-CoA as a starter molecule. These assays employed MCS and its necessary substrates instead of using malonyl-CoA (see Section 4.2.9). Reaction products were analyzed by LCMS with selected ion monitoring in ES^+ mode at m/z values of 153.19 corresponding to 5-propylbenzene-1,3-diol (divarin; an olivetol propyl analog), 155.16 corresponding to 4-hydroxy-6-propyl-pyran-2-one (butyryl diacetic acid lactone (BDAL); a PDAL analog), and 197.20 corresponding to both 2-hydroxy-6-(2-oxopentyl)pyran-4-one (butyryl triacetic acid lactone (BTAL); an HTAL analog), and 2,4-dihydroxy-6-propylbenzoic acid (divarinic acid; an OA analog) (Figure 4.21). When TKS was assayed in the absence of OAC, compounds were detected at 7.4 min (BDAL), 9.0 min (BTAL), and at 16.9 min (divarin). When TKS and OAC were assayed together, in addition to those peaks previously mentioned, another peak at 18.0 min was detected which corresponded exactly with SIR at 197.20 m/z (divarinic acid). An unknown compound eluting at 16.1 min was also detected by PDA and SIR at m/z 153.19; this peak is not seen when TKS was assayed alone.

4.3.10.4 With Octanoyl-CoA

TKS was assayed in the presence and absence of OAC using octanoyl-CoA as a starter molecule. These assays also used MCS and its required components instead of malonyl-CoA (See Section 4.2.9). Reaction products were analyzed by LCMS using selected ion monitoring in ES^+ mode at m/z values of 209.30 corresponding to 5-heptylbenzene-1,3-diol (heptylresorcinol; an olivetol heptyl analog), 211.27 corresponding to 6-heptyl-4-hydroxypyran-2-one (heptyl diacetic acid lactone (HDAL); a PDAL analog), and 253.31 corresponding to both 4-hydroxy-6-(2-oxononyl)pyran-2-one (octanoyl triacetic acid lactone (OTAL); an HTAL analog), and 2-heptyl-

4,6-dihydroxybenzoic acid (heptylresorcinolic acid; an OA heptyl analog) (Figure 4.22). When TKS was assayed in the absence of OAC, compounds were detected at 6.7 min (HDAL), 7.8 min (OTAL), 9.5 min (unknown), and at 10.6 min (heptylresorcinol). The unknown compound observed during PDA analysis eluting at 9.5 min is not detectable by SIR at 253.31; this compound is not heptylresorcinolic acid. The same compounds were detected when TKS and OAC were assayed, in addition to another compound, seen as a larger peak eluting at 9.4 min during PDA analysis; this compound (heptylresorcinolic acid) is clearly detectable by SIR at m/z 253.31 (Figure 4.22).

4.3.10.5 With Acetyl-CoA

TKS was assayed in the presence and absence of OAC using acetyl-CoA as a starter molecule. These assays used MCS and its required components instead of malonyl-CoA (Section 4.2.9). Reaction products were analyzed by LCMS using selected ion monitoring in ES^+ mode at m/z values of 125.14 corresponding to 5-methyl-1,3-benzenediol (orcinol; an olivetol methyl analog), 127.11 corresponding to 6-methyl-4-hydroxypyran-2-one (triacetic acid lactone (TAL); a PDAL analog), and 169.15 corresponding to both 2-acetonyl-6-hydroxy-pyran-4-one (tetra acetic acid lactone (TTAL); an HTAL analog), and 2,4-dihydroxy-6-methylbenzoic acid (orsellinic acid; an OA methyl analog) (Figure 4.23). Whether TKS was assayed with or without OAC, the assumed compounds were detected at 4.0 min (TAL), 4.8 min (TTAL), 8.1 min (orcinol), and at 10.1 min (orsellinic acid). An unknown compound was also observed during PDA analysis at 270 nm eluting at 3.5 min, which is also detectable by SIR at m/z 169.15 (Figure 4.23). There is little difference between assays conducted with or without OAC suggesting that shorter alkyl tetraketides may spontaneously cyclize to orsellinic acid.

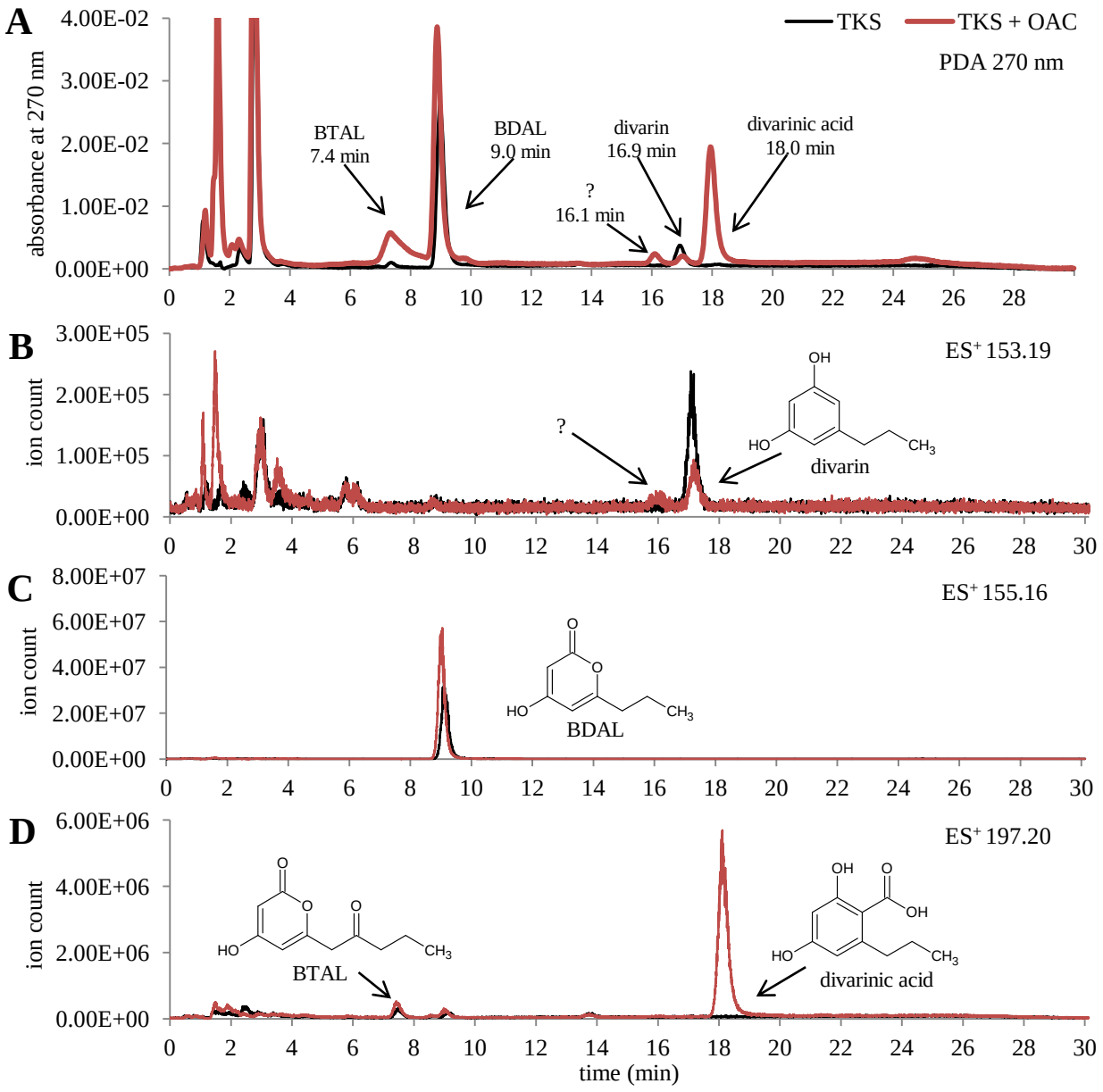


Figure 4.21 LCMS of TKS/OAC assays conducted with butyryl-CoA.

LCMS chromatograms from TKS assays performed in the presence (red) and absence (black) of OAC are overlaid. (A) LC traces align well with selected ion monitoring in ES^+ mode capable of detecting (B) divarin (16.9 min; m/z 153.19), (C) BDAL (9.0 min; m/z 155.16), (D) BTAL (7.4 min; m/z 197.20) and divarinic acid (18.0 min; m/z 197.20). An unknown compound was also detected using PDA monitoring at 16.1 min (A) and SIR monitoring at m/z of 153.19 (B). No divarinic acid was detected when TKS was assayed in the absence of OAC.

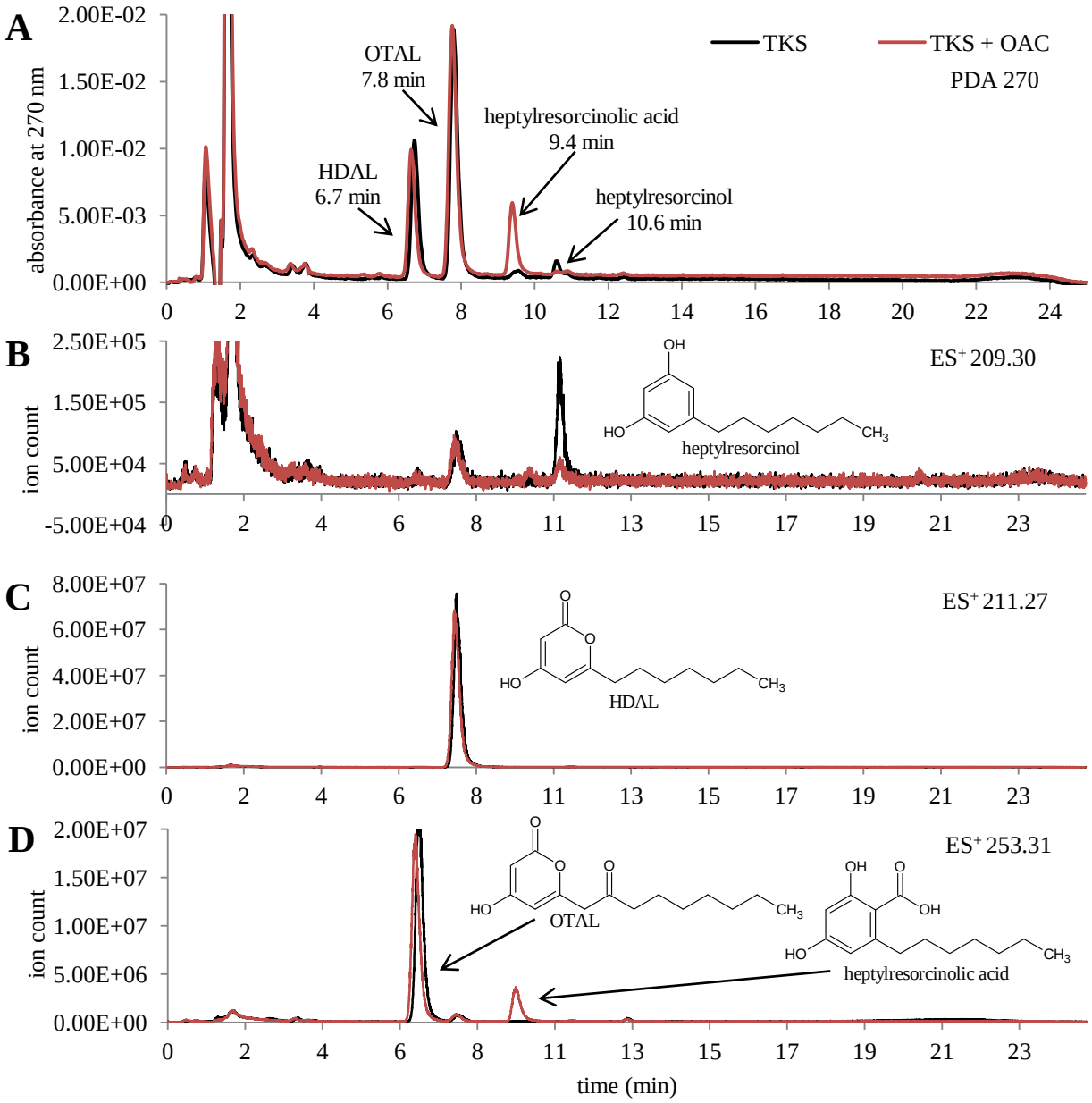


Figure 4.22 LCMS of TKS/OAC assays conducted with octanoyl-CoA.

LCMS chromatograms from TKS assays performed in the presence (red) and absence (black) of OAC are overlaid. (A) LC traces align well with selected ion monitoring in ES^+ mode capable of detecting (B) heptylresorcinol (10.6 min; m/z 209.30), (C) HDAL (7.8 min; m/z 211.27), (D) OTAL (6.7 min; m/z 253.31) and heptylresorcinolic acid (9.4 min; m/z 253.31). The detection of heptylresorcinolic acid is confirmed by its appearance at 9.4 min during PDA analysis (A) and its corresponding peak during SIR at m/z 253.31 (D), which was not detected when TKS was assayed in the absence of OAC.

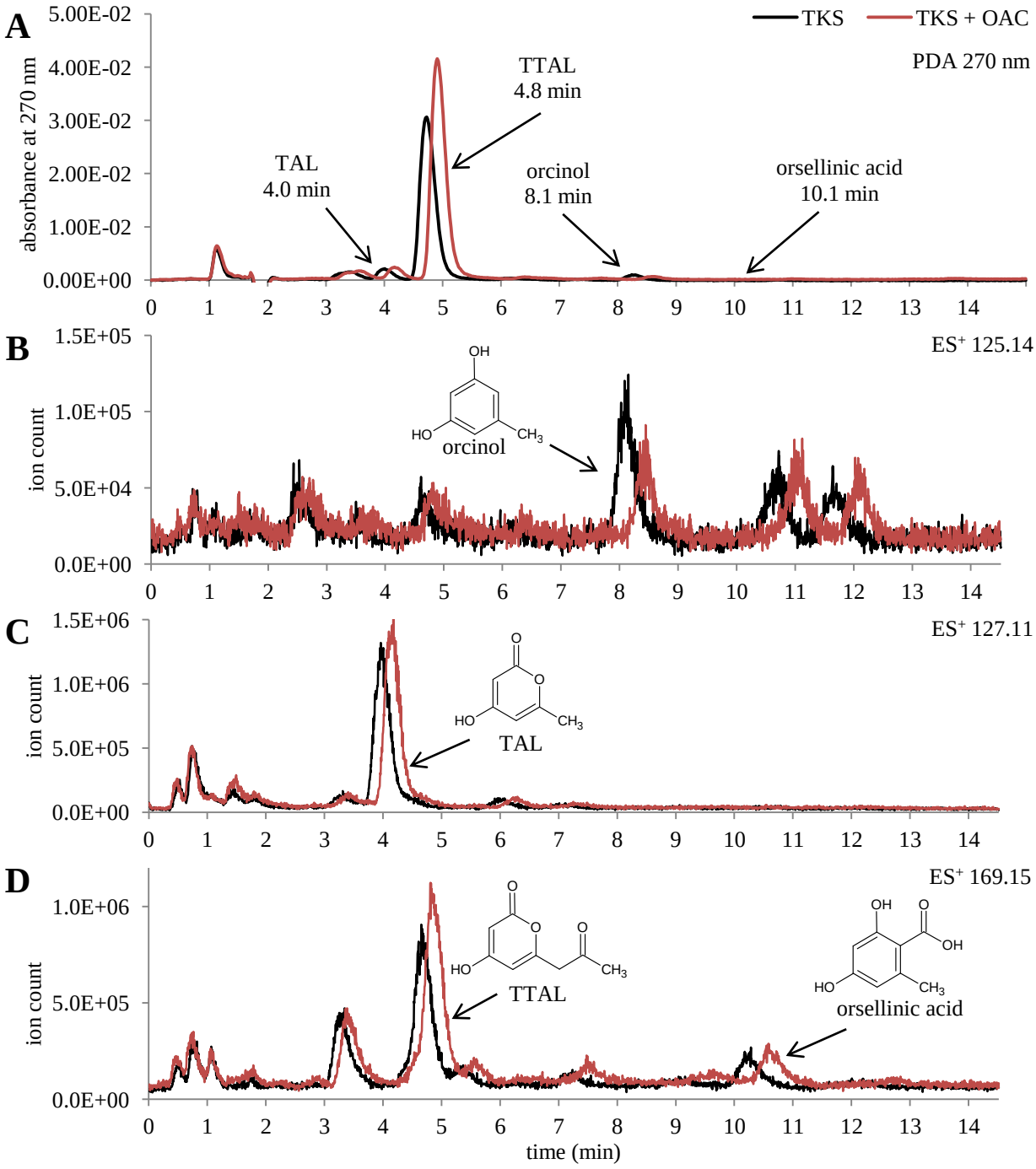


Figure 4.23 LCMS of TKS/OAC assays conducted with acetyl-CoA.

LCMS chromatograms from TKS assays performed in the presence (red) and absence (black) of OAC. (A) LC and MS ES^+ results are aligned and capable of detecting what is assumed to be (B) orcinol (8.1 min; m/z 125.14), (C) TAL (4.0 min; m/z 127.11), (D) TTAL (4.8 min; m/z 169.15) and orsellinic acid (10.1 min; m/z 169.15), which were detected in all assays performed.

4.4 Discussion

TKS assays conducted in the presence of candidate polyketide cyclases led to the discovery of OAC, an enzyme capable of modifying TKS product formation resulting in the synthesis of OA (Gagne et al., 2012). This discovery warranted the biochemical characterization of OAC, the main focus of this chapter.

4.4.1 Phylogenetic Relationship of OAC

Blast searches for OAC homologs indicate that similar proteins are ubiquitous in plants. Their strong conservation indicates a beneficial property in plants, probably related to salt and heat stress responses, as their family name suggests (*i.e.* SRABBs). The occurrence of similar sequences to OAC in bacterial genomes (namely *Rhizobium*) is an indicator that they may originate from ancestral prokaryotes, and may have been introduced to plants by early horizontal gene transfer events, as it has been hypothesized to occur in *Nicotiana* (Intrieri and Buiatti, 2001). Phylogenetic analysis divides the OAC homologs into two specific clades suggesting that a primary OAC homolog (clade A) is conserved, and may have duplicated so to form a clade of paralogs (clade B). The relationship OAC shares with these clades A and B appears remote, having formed its own branch with clade A, but includes homologs from *Mimulus guttatus* (common monkey-flower), and *Populus trichocharpa* (poplar).

4.4.2 The Optimum TKS/OAC Ratio

Ratio experiments were designed to determine the amount of OAC required for maximal OA production. The amount of enzyme used in these assays reported values in pmol instead of µg values to determine if a stoichiometric relationship existed between TKS and OAC; results suggest that no such relationship exists and that maximal OA production occurred when

OAC:TKS ratios were approximately 40:1. The largest amount of OA production was observed when 140 µg of OAC (*i.e.* 7,500 pmol) was assayed with 9 µg of TKS (200 pmol). Larger amounts (>7,500 pmol) of OAC did not increase OA production, indicating that peak OAC activity had been achieved, or that self-inhibition of OAC (and/or possibly TKS inhibition) was occurring due to its increased concentration. Based on its structural classification, it is presumed that OAC exists as a dimer in solution and that at greater concentrations a higher ordered oligomer could exist, as it does with the OAC homolog, *Populus tremula* SP1, which is capable of forming dodecameric quaternary structures (Dgany et al., 2004). This potential oligomerization could lead to structural changes that result in self-inhibition, which is known to occur in other enzymes such as 12-oxophytodienoate reductase from *Lycopersicon esculentum* (tomato), which exhibit a weak tendency for dimerization correlating with a loss of activity (Breithaupt et al., 2006). Results themselves do not suggest a loss of activity, and may instead indicate a maximum limit of products being reached, and that the effects seen when >7500 pmol OAC were used is simply a leveling-off of OA production. Results suggest that 100 pmol of TKS should be assayed with 3,500 pmol of OAC for optimum OA production, but further analyses were conducted using a TKS:OAC ratio of 100:1,000 pmol due to the huge amounts of recombinant OAC that would otherwise be required (where 1,000 pmol of OAC is approximately 19 µg of enzyme).

4.4.3 The Optimum Temperature for *In Vitro* OA Production

Assays designed to determine the optimum temperature for OA production gave insights about the TKS/OAC reaction. The results from TKS assays conducted in the absence of OAC indicate that maximal product formation occurred at ca. 40°C. When TKS was assayed in the presence of

OAC, the largest amount of OA was produced at a lower temperature of 15°C, and OA production was reduced at temperatures exceeding 25°C. These results suggest that an unstable intermediate is produced by TKS, which undergoes a spontaneous decarboxylative cyclization at higher temperatures resulting in the formation of olivetol. Conversely, at lower temperatures the tetraketide intermediate is stabilized, allowing it to reach OAC, where it is cyclized to OA before any thermo-dependent spontaneous cyclization occurs. Based on these data, future assays optimized for OA production were conducted at 20°C.

4.4.4 The Optimum Buffer and pH for OAC

The choice of buffering agents also played a key role in determining the optimum for OA production. Results from combinatorial assays performed in the presence of sodium phosphate, PIPES, HEPES, TRIS, and citrate were directly compared and results indicate that HEPES is the best buffering system for OA production. Furthermore, a lower concentration of HEPES was deemed beneficial for tetraketide production, and a 20 mM HEPES buffering solution was used in subsequent assays. The activity of OAC was also compared over a wide pH range using a number of buffers due to the limited effective pH range of HEPES. Combined results indicate that maximal OAC activity occurs at a pH of 6.4, but this pH is not possible with HEPES. Nevertheless, the amounts of OA detected from assays conducted in HEPES buffer at pH 7.0 were greater than those amounting when assays were conducted in phosphate, TRIS, or PIPES at pH 6.4. It was decided that future assays would use HEPES at pH 7.0. At a later date, citrate and MOPSO buffer systems were also compared to HEPES, and showed that MOPSO was as good as HEPES as a buffering agent, if not superior. In retrospect, the use of MOPSO would have been beneficial during all work that continued, since its pH range is 6.2 to 7.6, and an optimum

pH of 6.4 could have been achieved. However, the MOPSO/HEPES buffer comparison was tested after many assays had already been performed, and in the interest of time and resources, the use of HEPES as a buffer continued throughout this study. The cause for the differences in TKS and OAC activity in the various buffer systems is not known.

4.4.5 OAC Does Not Require a Metal for Activity

The TKS/OAC reaction was assayed in the presence of metals to determine if they would affect TKS and OAC activity, potentially leading to an augmentation in OA production. Additionally, the possibility that OAC employs a type II aldol mechanism to achieve OA cyclization, which would require the participation of a metal ion, was also explored (Silverman, 2002). It was also reasonable to suspect that OAC required a metal, since Mg^{2+} has been identified in the AtHS1 crystal structure (Bingman et al., 2004). Results clearly indicate that no metal tested could increase OAC activity. The effect that some transition metals have on the TKS/OAC product profile is difficult to explain without further experimentation, but the observation that Fe^{2+}, Co^{2+}, Zn^{2+} and Cu^{2+} affect polyketide chemoselectivity is intriguing enough to warrant a directed study in the future. It is easy enough to explain a loss of activity when OAC is assayed in the presence of these metals, but the amplification of olivetol production to levels that are three times the amounts observed in assays conducted in the absence of metals, coupled to a reduction of PDAL derailment products suggests a possible route to stabilizing TKS, and putatively other type III PKS *in vitro* reactions. These findings indicate that there is no metal requirement for OAC. Future experiments that would examine this chemoselectivity phenomenon should include a broader spectrum of metals to test and vary the concentrations of these metals when being assayed with TKS.

In support of the notion that neither TKS nor OAC requires any metal for their activity are the results from assays conducted in the presence of EDTA. The results showed that, although TKS product accumulation is decreased when TKS and OAC are assayed with EDTA, enzyme activity is not abolished, and OA accumulation is hardly affected (Figure 4.17). The findings discount the metal-catalyzed aldol condensation as a plausible mechanism for OAC. Likewise, no other bacterial cyclase involved with type II PKSs has been reported to require a metal (Schneider, 2005).

The inclusion of EDTA also caused an increase in PDAL production, suggesting EDTA destabilizes the TKS reaction by causing the premature release of nascent polyketide intermediates, although notably, OAC activity remained unaffected. The effects caused by EDTA and metals may be due to direct interaction on the enzymes themselves, but may also have been caused by their reactivity to a yet unidentified labile polyketide intermediate that remains in solution until it cyclizes.

4.4.6 The use of MCS in assays abolishes the need for malonyl-CoA

Assays that utilized MCS bypassed the requirement of malonyl-CoA in TKS/OAC assays, but required the inclusion of sodium malonate, ATP, Mg^{2+}, and free CoA instead. These assays were primarily conducted to determine if MCS could augment OA production during TKS/OAC assays. Assays performed in the presence of MCS were capable of successfully generating malonyl-CoA required by TKS, although a reduction in product accumulation was observed. This reduction in overall TKS/OAC activity could be due to the addition of Mg^{2+} and/or free CoA, which have each been shown to be destabilizing agents. It should be noted that the optimization of the MCS reaction with TKS and OAC has not been conducted, and that future

work employing MCS could benefit if this were to take place, which would most likely lead to an increase in product yield. Nevertheless, the inclusion of MCS in the TKS/OAC reaction was successful in that these reactions could be conducted without exogenous malonyl-CoA. Assays were also conducted with free CoA to determine how well this compound could be tolerated. As previously mentioned, a decrease in the overall TKS/OAC reaction output is correlated to the concentrations of free CoA in enzyme assays, but the concentration of free CoA used in later analyses (0.20 mM CoA) did not interfere with the quantification of assay products.

4.4.7 Neither HTAL, PDAL, nor olivetol are substrates for OAC

The possibility that OAC acts directly on TKS products was explored by assaying OAC in the presence of HTAL, PDAL, and olivetol, but no significant OA production was detected. Trace amounts of OA were detectable when 1 mM HTAL was incubated with OAC, although this small amount was almost negligible (20 pmol), since typically TKS reaction generates much smaller amount of HTAL than the amounts used in these assays. If OAC was capable of opening and re-cyclizing HTAL into OA, it would be anticipated that OA accumulation would be comparable to a typical TKS/OAC reaction, where OA accumulated to approximately 500 pmol. It is noted that OA was detected when OAC was assayed with 1 mM HTAL, but that the minute amount may have been produced following spontaneous HTAL ring opening followed by spontaneous or possibly OAC-catalyzed cyclization to form OA. The spontaneous opening and closing of lactones which results in the formation of resorcylates has been previously reported with starter-chain homologs (*i.e.* TTAL) (Guilford et al., 1968). I propose that a similar reaction also occurs, although at a slower rate, with HTAL. The potential involvement of TKS during HTAL ring opening was also assessed by including TKS during OAC assays performed in the

presence of HTAL, but no OA was detected, conclusively discounting this possibility. Similarly, the putative involvement of free CoA as a cofactor was also investigated during OAC/HTAL assays, but no OA was detected during LCMS analysis.

4.4.8 Polyketide Analogs are Cyclized to their Respective Resorcylic Acid by OAC

TKS/OAC assays were also performed using alternative CoA-esters besides hexanoyl-CoA, which included butyryl-CoA, octanoyl-CoA, and acetyl-CoA. Although no standards were used, the detection of polyketide analogs was made possible by comparing product UV spectra produced in these assays with those from standard TKS/OAC assay products that used hexanoyl-CoA as a starter. Additionally, the use of LCMS SIR for masses of the theoretical products was also capable of confirming the identities of the polyketide analogs produced (Figures 4.21, 4.22, and 4.23). Acetyl-CoA, butyryl-CoA and octanoyl-CoA were each accepted as starter compounds by TKS, and appeared to be successfully incorporated into triketide and tetraketide products, which included PDAL, HTAL, and olivetol analogs. In addition to these analogs, TKS assays conducted in the presence of OAC were also capable of producing divarinic acid and heptylresorcylic acid, when butyryl-CoA and octanoyl-CoA were used as alternative starter compounds, respectively. These results contradict previous reports that TKS cannot produce HTAL analogs when either butyryl-CoA or octanoyl-CoA replaces hexanoyl-CoA as starter CoA ester, and that TKS cannot produce heptylresorcylic acid when octanoyl-CoA is used (Taura et al., 2009). The lack of polyketide analogs in assays conducted by others may be due to differences in buffering system, the lower pH employed, or subtle differences between TKS and OLS.

Findings also show that TKS is capable of using acetyl-CoA as a starter molecule, and results in the production of TAL, TTAL, orcinol, and orsellinic acid (Figure 4.23). These assays differ from assays using butyryl-CoA or octanoyl-CoA, in that TKS assays that use acetyl-CoA are capable of accumulating a resorcylic acid analog (orsellinic acid in this case) whether they are performed in the presence or absence of OAC. When butyryl-CoA or octanoyl-CoA were used as priming agents, the accumulation of their resorcylic acid analogs was only produced when OAC was present. This difference in the cyclization can be attributed to the variability in length and bulkiness of the primer's alkyl groups, and the propensity of the synthesized tetraketide's lactone ring to open and close in solution. The ability of TTAL to spontaneously and easily reform into orcinol and orsellinic acid has been previously reported (Bentley and Zwitkowits, 1967; Guilford et al., 1968). Guilford et al. (1968) have also demonstrated that phenacylpyrone, a TTAL lactone analog attached to a bulky phenyl group, is a moderately stable lactone capable of withstanding ring opening and subsequent conversion to its recorcylate/resorcinol analog, whereas TTAL, which contains only a methyl group instead of a phenyl ring, can open easily. Therefore, it is not surprising to discover orsellinic acid is a product of TKS assays that use acetyl-CoA when conducted in the absence of OAC.

The ability of TKS and OAC to synthesize different resorcylic acids by using different starting molecules is in agreement with the occurrence of cannabinoid analogs observed in nature. These analogs replace the otherwise normal pentyl side chain of cannabinoids with methyl (Vreet et al., 1972), propyl (Merkus, 1971), butyl (Harvey, 1976), and possibly even heptyl side chains (Isbell, 1973). The biosynthesis of these cannabinoid analogs would proceed through the prenylation of short-chain (1-8 carbon) resorcylic acids which would have been

generated by TKS and OAC. The plasticity observed in TKS is reminiscent of GCHS2, a non-CHS type III PKS from *Gerbera hybrida* also capable of accepting acetyl-CoA as a starter molecule (Eckermann et al., 1998). Unlike the traditional CHSs that synthesize naringenin chalcone from phenylpropanoid CoA esters, GCHS2 is a pyrone synthase that synthesizes TAL after two condensations with acetyl-CoA. GCHS2 is also capable of using only malonyl-CoA as a primer following decarboxylation to acetyl-CoA. Similar activity was not observed with TKS.

5. ANALYSIS OF PHYSICAL INTERACTION BETWEEN TKS AND OAC

5.1 Introduction

This chapter focuses on experiments designed to examine the physical relationship between TKS and OAC, which may also provide insight into the mechanism of OA biosynthesis. Despite the biochemical evidence presented in Chapter 4, showing that TKS and OAC function together to produce OA, fundamental questions about to the mechanism of OA biosynthesis still need to be addressed. One issue was whether these enzymes exist in the same cellular compartment, for if they do not, it would be difficult to suggest that they participate together in an enzymatic reaction. The possibility that TKS and OAC physically interact is reasonable, given that OAC resembles a type II cyclase, and that type II polyketide biosynthesis is accomplished by multienzyme complexes (Staunton and Weissman, 2001). Multienzyme complexes (metabolons) have also been shown to exist in plants, as in Calvin cycle complexes (Suss et al., 1993), serine and cysteine biosynthesis enzyme complexes (Prabhu et al., 1996; Saito et al., 1995), and the oxidative pentose phosphate metabolon (Debnam et al., 1997). Winkel-Shirley (1999) suggests that metabolons frequently occur in the cell and constitute a structured organization of cellular components instead of loosely associated soluble enzymes (Winkel-Shirley, 1999). This makes sense, considering that cellular activities could be more tightly regulated in this way, as well as increasing metabolic efficiency, increasing local substrate concentrations, and partitioning metabolites to similar branching pathways. Earlier reports from Stafford (1974) also suggest that enzyme complexes exist so to facilitate the biosynthesis of products that are localized in specific subcellular compartments (Stafford, 1974). Interestingly, enzymes from both the

phenylpropanoid and flavonoid biosynthesis pathways have been shown to interact. Winkel-Shirley's (1999) review on enzyme complex association in the phenylpropanoid and flavonoid biosynthesis pathways suggests that: 1) metabolites are channelled from one enzyme to another as in the case of phenylalanine ammonia-lyase (PAL) and cinnamate 4-hydroxylase (C4H), where ^{3}H-phenylalanine is preferentially used as a substrate instead of ^{14}C-cinnamic acid in the production of labelled *p*-coumaric acid (Czichi and Kindl, 1977); 2) enzymes are co-localized at microsomal and endoplasmic reticulum sites, as is evident by cell fractionation experiments conducted by Grisebach and Fritsch (1975), and similarly Hrazdina et al. (1978) who showed that CHS pellets with the microsomal fraction of *Tulipa* and *Hippeastrum* tissues, and later by Wagner and Hrazdina (1984), who demonstrated that PAL, C4H, and CHS all co-fractionated with NADPH:Cyt *c* reductase, an endoplasmic reticulum marker; 3) enzymes involved in the phenylpropanoid/flavonoid biosynthesis pathway are induced in a co-ordinate manner, which has led to the classification of early and late genes associated with the synthesis of flavonoids, and the later synthesis of proanthocyanidins, anthocyanidins and flavonols (Pelletier et al., 1997); 4) direct protein-protein interactions do occur, as reported by Burbulis et al. (1999), who used a yeast two-hybrid (Y2H) system to show direct interaction between *A. thaliana* CHS, chalcone isomerase (CHI) and dihydroflavonol 4-reductase (D4R), and used polyclonal antibodies generated against CHI to co-immunoprecipitate CHS and D4R.

To determine if TKS and OAC co-localize *in vivo*, the subcellular localization of fluorescent tagged TKS and OAC was performed. This allowed the visualization of these proteins *in planta* using confocal microscopy following agroinfiltration in *Nicotiana benthamiana* leaves. Once the co-localization of the enzymes was confirmed, a Y2H assay was

performed to determine if TKS and OAC physically interact. If a physical interaction existed, it would imply that TKS physically directs a transient intermediate to OAC, as in the case of PAL and C4H (Czichi and Kindl, 1977), or possibly that TKS activity is modified by OAC through an allosteric mechanism that alters the TKS active site.

Further experiments attempted to determine if TKS produces and subsequently releases a polyketide intermediate, which is then bound and cyclized by OAC. TKS/OAC reactions were conducted by exploiting the poly-histidine tag in recombinant TKS using magnetic Talon beads, allowing TKS to be removed during enzyme assays and then observing whether or not OA could be produced in the absence of TKS.

Similar assays, designed to assess the enzymes' ability to function independently, were performed using a microdialysis apparatus which separated TKS and OAC. These experiments have provided important insights into the physical relationship between TKS and OAC, as well as providing valuable evidence that supports the notion of a linear polyketide intermediate being produced and released from TKS before it is acquired and cyclized by OAC.

5.2 Materials and Methods

5.2.1 Subcellular Localization of TKS and OAC

The TKS-cyan fluorescent fusion protein (TKS:CFP) and OAC-yellow fluorescent fusion protein (OAC:YFP) constructs were generated by amplifying TKS and OAC without C-terminal stop codons using Pfu Ultra II DNA Polymerase (Stratagene) and the primers and cycling conditions shown in Tables 5.1 and 5.2. PCR products were A-tailed by incubation at 72°C for 15 min along with 50 U of Taq polymerase (Invitrogen) and 0.10 mM dATP before purification using a Minelute PCR kit (Qiagen). PCR products were ligated into the pCR8/GW/topo entry vector (Invitrogen), resulting in the pSG054 and pSG055 plasmids harboring TKS and OAC without stop codons, respectively, that were separately used to transform *E. coli* DH5α competent cells (Invitrogen). Plasmids were isolated using a Qiaprep Spin Miniprep Kit (Qiagen) from DH5α cultures after they were grown in LB in the presence of 50 mg/mL spectinomycin overnight at 37°C. Following sequence verification (PBI DNA Technologies Unit), pSG054 was subcloned into the pEARLEY102 (Invitrogen), resulting in the TKS:CFP construct. Similarly, pSG055 was subcloned into pEARLEY101 (Invitrogen), creating the OAC:YFP construct. Following LR recombination, constructs were sequenced to confirm orientation using the primers described in Table 5.2.

Expression constructs were transformed by electoporation in competent *Agrobacterium tumefaciens* LBA4404 and plated onto selective LB in the presence of 10 µg/mL of rifampacin and 50 µg/mL of kanamycin (Weigel and Glazebrook, 2006). Freshly cultured transformants were grown to an OD_{600} of 0.2 and used to agroinfiltrate two-week old *Nicotiana benthamiana* leaves (Sparkes et al., 2006). Following a two-day incubation period, plant tissue near

agroinfiltration sites were cut into 1 x 1 cm squares and placed in $_dH_2O$ on a microscope slide prior to being examined by confocal microscopy. A Zeiss LSM510 confocal microscope was used to visualize the TKS:CFP construct after excitation with a diode laser set to 458 nm and to capture images using a 475-525 nm bandpass filter. The OAC:YFP construct was visualized by setting the diode laser to 514 nm and using a 530-600 nm bandpass filter. Images were processed using the Zen software suite (Carl Zeiss Microscopy).

Table 5.1 Primers used during the creation of TKS:CFP and OAC:YFP

Name	Forward Primer	Reverse Primer
TKS:CFP	ATGAATCATCTTCGTGCTGAG	ATATTTGATGGGAACACTACG
OAC:YFP	ATGGCAGTGAAGCATTTGATT	CTACTTTCGTGGTGTGTAGTC
TKS:CFP sequencing	ATGAATCATCTTCGTGCTGAG	CTTTCGTGGTGTGTAGTC
OAC:YFP sequencing	ATGGCAGTGAAGCATTTGATT	CTTTCGTGGTGTGTAGTC

Table 5.2 Details regarding the creation TKS:CFP and OAC:YFP constructs

Name	Template	Melting[1]	Annealing[1]	Elongation[1]	Destination	Selection[2]
TKS:CFP	pHIS8-TKS	95°C/20s	50°C/20s	72°C/18s	pEARLEY102	Kan^R
OAC:YFP	pET100-OAC	98°C/20s	55°C/20s	72°C/18s	pEARLEY101	Kan^R

1 - The reactions were cycled 30-35 times before ending with 7 min at 72°C.
2 - Kanamycin resistance (Kan^R).

5.2.2 Yeast Two-Hybrid Methods

Experiments used a ProQuest Two-Hybrid System (Invitrogen). The system uses two specialized Gateway plasmids, pDest32 and pDest22, for rapid cloning, expression and analysis of proteins of interest. The pDest32 plasmid contains a GAL4 DNA-binding domain (DBD), one of two domains derived from the GAL4 transcription factor, and is responsible for binding an upstream operator sequence. Proteins of interest fused to this domain are collectively known as bait protein. The activating domain (AD) is the other half of the GAL4 transcription factor and is harbored by the pDest22 plasmid. Proteins of interest fused to this domain are referred to as prey proteins. β-galactosidase assays of fusion proteins allowed for blue/white screening using 5-bromo-4-chloro-3-indolyl- beta-D-galactopyranoside (X-gal) as a substrate. If an interaction existed between bait and prey fusion proteins, X-gal would be cleaved, resulting in an insoluble blue 5, 5'-dibromo-4,4'-dichloro-indigo which acts as an colorimetric indicator.

A number of plasmids were required for these experiments, necessitating the cloning of TKS and OAC into the pDest32 and pDest22 vectors. Furthermore, a combination of plasmids were used to transform the *Saccharomyces cerevisiae* strain MaV203, resulting in a collection of different transformants (results Section 5.3.2). Plasmids used in this work are summarized in Table 5.3.

Table 5.3 Plasmids used to transform MaV203 yeast used in Y2H experiments

Name	Function	Derived from	Selection[1]
pDEST32	Bait vector	n/a	LEU2 and Gen^R
pDEST22	Prey vector	n/a	TRP1 and Amp^R
pEXP32/Krev1	Control bait	pDEST32	LEU2 and Gen^R
pEXP22/RalGDS-wt	Strong positive control	pDEST22	TRP1 and Amp^R
pEXP22/RalGDS-m1	Intermediate positive control	pDEST22	TRP1 and Amp^R
pEXP22/RalGDS-m2	Negative control	pDEST22	TRP1 and Amp^R
pRT030	TKS (entry vector)	pCR8/GW/TOPO	Spc^R
pSC033	OAC (entry vector)	pCR8/GW/TOPO	Spc^R
pSG057	TKS in prey vector	pDEST22 + pRT030	TRP1 and Amp^R
pSG058	TKS in bait vector	pDEST32 + pRT030	LEU2 and Gen^R
pSG059	OAC in prey vector	pDEST22 + pSC033	TRP1 and Amp^R
pSG060	OAC in bait vector	pDEST32 + pSC033	LEU2 and Gen^R

1 – Antibiotic resistance to gentamycin (Gen^R), ampicillin (Amp^R), or spectinomycin (Spc^R).

5.2.3 Plasmid Construction

Both Gateway-compatible TKS (pSC033) and OAC (pRT030) were subcloned into each pDEST22 and pDEST32 by LR recombination as directed by the manufacturer's protocol. Reagents were supplied as part of the ProQuest Two-Hybrid System (Invitrogen). The reactions generated the pSG057, pSG058, pSG059, and pSG060 constructs (Table 5.3). Plasmids were transformed into *E. coli* Mach 1 cells (Invitrogen) and, following their culturing, were isolated using a QIAprep Spin Miniprep Kit (Qiagen) before being sequenced using the recommended ProQuest kit primers (Invitrogen) (Table 5.4).

Table 5.4 Sequencing primers for pDEST22 and pDEST32 based vectors

Function	Sequence
pDEST32 forward sequencing primer (23mer)	AACCGAAGTGCGCCAAGTGTCTG
pDEST22 forward sequencing primer (21mer)	TATAACGCGTTTGGAATCACT
pDEST32/22 reverse sequencing primer (23mer)	AGCCGACAACCTTGATTGGAGAC

5.2.4 Yeast Transformants

Once sequences were confirmed, *Saccharomyces cerevisiae* MaV203 competent cells were transformed with the generated plasmids (Table 5.3). MaV203 cells were made competent by initially inoculating 10 mL of liquid YPAD media (1% Bacto-yeast extract, 2% Bacto-peptone, 2% glucose, 0.01% adenine sulfate [pH 6]) and culturing overnight at 30°C. The overnight culture was diluted to an OD_{600} of 0.4 in 50 mL of YPAD and incubated for an additional 4 h. The culture was centrifuged at 1,000 g and resuspended in 40 mL of sterile TE (10 mM TRIS-HCl [pH 7.5], 1 mM EDTA). Cells were again centrifuged at 1,000 g (supernatant discarded) and resuspended in 2 mL of sterile 1 x LiAc/0.5x TE (100 mM LiAc, 5 mM TRIS-HCl [pH 7]). Cells were incubated in this solution for 10 min before proceeding to transformations. A total of 1 µg of DNA was used in each transformation, where 500 ng of DNA from each plasmid was used. An example of this is SG100, where 1 µL of 500 ng/µL pEXP32/Krev1 was used with 1 µL of 500 ng/µL pEXP22/RalGDS-*wt* resulting in 1 µg of total DNA being used in this transformation. For each transformation 1 µg of DNA was added with 100 µg denatured sheared salmon sperm DNA to 100 µL of competent yeast cells. 700 µL of 1 x LiAc/40% PEG-3350/1x TE was then added and mixed. Reactions were incubated for 30 min at 30°C, followed by the addition of 88 µL DMSO. Cells were then heat-shocked at 42°C for 7 min, followed by a 1,000 g centrifugation for 10 s. Supernatant was removed, and cells were resuspended in 1 mL of 1 x TE. Cells were again pelleted at 1,000 g for 10 s and resuspended in 100 µL TE, before being plated onto selective SC-Leu-Trp plates (0.67% Bacto-yeast nitrogen base without amino acids (Sigma), 2% glucose (w/v), 2% Bacto agar (w/v), 0.2% dropout mix (w/v) containing equal (w/w) proportions of adenine sulfate, alanine, arginine, aspartic acid, asparagine, cysteine, glutamic acid,

glutamine, glycine, isoleucine, lysine, methionine, phenylalanine, proline, serine, threonine, tyrosine, and valine [pH 5.9]). Plates were incubated for three days at 30°C. After three days colonies were selected and restreaked onto fresh SC-Leu-Trp plates and incubated for another 3 days at 30°C. A negative control lacking plasmid was also set up, although no cells were found to grow.

5.2.5 Confirmation of DNA Inserts

Yeast DNA was prepared using a QIAprep Spin Miniprep Kit (Qiagen), and a protocol adapted by Jones (2000). MaV203 yeast transformants SG103, SG200, SG201, SG202, and SG203 were cultured for two days at 30°C. Cells were pelleted (5 min at 5,000 g) and resuspended in 250 mL P1 buffer containing 0.1 mg/mL RNase A. Approximately 100 mL of acid-washed glass beads (Sigma) were used to rupture cells over 2 min in a minibeadbeater (Biospec). The remainder of the DNA extraction was conducted as directed by Jones (2000), although 25 mL of elution buffer (EB) was used in the final step instead of 50 mL. A 2 µL aliquot of purified DNA was used to individually transform 50 mL of *E. coli* Top 10' electrocompetent cells (Invitrogen), which were plated onto LB + 100 µg/mL carbenicillin and LB + 10 µg/mL gentamycin and grown overnight at 37°C. Cultures were successfully transformed, except for SG103 and SG 200 which did not grow on LB + 10 µg/mL gentamycin. Single colonies were used to inoculate 5 mL liquid LB + 10 µg/mL gentamycin or LB + 100 µg/mL carbenicillin and grown overnight at 37°C. DNA was purified using a QIAprep Spin Miniprep Kit (Qiagen), quantified using a NanoDrop 8000 Spectrophotometer (Thermo Scientific) and sequenced using ProQuest-recommended sequencing primers (Invitrogen) (Table 5.4). Although SG200 did not grow on LB + 10 µg/mL

gentamycin, the TKS cassette was detected by direct sequencing using either the pDEST32 or pDEST22 forward sequencing primers.

5.2.5.1 Determination of 3-AT Concentration used in Histidine Dropout Media

Before proceeding with the full characterization of transformants, the amounts of 3-aminotriazole (3-AT) needed to inhibit HIS3 self-activation were assessed. Yeast transformants SG100, SG101, SG102, SG200, SG201, and SG202 were grown overnight in 5 mL liquid SC-Leu-Trp medium at 30°C. Cells were pelleted and resuspended in 1 mL sterile dH_2O. Cells were again pelleted and resuspended in 1 mL sterile water. Cultures were diluted with sterile water until the OD_{600} of a 10x dilution (*i.e.* 100 µL of culture + 900 µL water) was 0.1 ± 0.01. This allowed the standardization of cultures with an estimated starting OD_{600} of 1.0 ± 10%. These cultures were serially diluted resulting in OD_{600} concentrations of 0.5, 0.25, 0.125, 0.0625, and 0.03125 (based on original OD_{600}). A 5 µL aliquot from each of these dilutions was spotted onto several SC-Leu-Trp-His plates containing either 10, 25, 50, or 100 mM 3-AT, and incubated at 30°C for three days.

5.2.5.2 Characterization of Transformants

Two-day old starter cells that had been grown at 30°C were pelleted (5,000 g for 10 min), the supernatant was removed, and cells washed with 2 mL sterile H_2O (this step removed excess amino acids). The cells were again pelleted and resuspended into 1 mL sterile water then diluted until an OD_{600} of 1.0 ± 0.1 was achieved. These cells were used to spot 5 µL aliquots of 1x, 2x and 10x dilutions onto SC –Leu –Trp –His + 50 mM 3-AT and onto YPAD plates covered with an 82 mm diameter nylon membrane specialized for colony and plaque hybridization (Roche). The SC –Leu –Trp –His + 3-AT plates were incubated for three days at 30°C and then observed

for colony formation. The YPAD + membrane plates were incubated for 24 h at 30°C. The nylon membrane was placed in liquid N_2 for 1 min. The membrane was then thawed on two saturated 150 mm diameter filter papers (Whatman) in a 150 mm petri dish. The filter papers were saturated with 8 mL X-gal solution (10 mg X-gal in 100 µL DMF + 60 µL 2-mercaptoethanol in 10 mL (filter sterile) Z-buffer (0.06 M Na_2HPO_4-$7H_2O$ + 0.04 M NaH_2PO_4-H_2O + 0.01 M KCl + 0.001 M $MgSO_4$-$7H_2O$ in H_2O [pH 7]). Petri dishes were covered and incubated 24 h at 37°C.

5.2.6 TKS Magnetic Pulldown Methods

The poly-histidine tag in the pHIS8-TKS construct was exploited using Talon magnetic beads (Clontech) to remove TKS from solutions containing malonyl-CoA and hexanoyl-CoA. Reaction mixtures were allowed to incubate with TKS for a given time, after which the TKS was removed, and the supernatant was transferred to a new reaction vessel containing OAC. The accumulation of OA was then analyzed by LCMS.

A new OAC construct that was codon-optimized for its expression in *E. coli* (pJE60174) was used during this work. The pJE60174 OAC construct was expressed, purified (as in Section 3.1.1), and digested in the presence of Thrombin to remove its poly-histidine tag. Protein solution eluting from Talon beads was concentrated to a 1 mL solution using YM10 centrifugal filter device (Millipore) and rinsed using a 5 mL solution of 20 mM HEPES, 50 mM NaCl, and 5 mM DTT before being concentrated again to 1 mL and transferred to a Spectra-Por Float-A-Lyzer with a 5 kDa MWCO (Sigma). A 10 U aliquot of high activity Thrombin (Millipore) was added to the dialysis tubing. The solution was allowed to digest and dialyze overnight at 4°C. The solution was removed from the dialysis tubing and incubated at 4°C for 1 h with 1 mL Talon (Clontech). The Talon was added to a gravity flow column and the wash-through was collected

and concentrated using a YM10 centrifugal filter device (Millipore). This solution contained the cleaved OAC without a His-tag.

Reaction mixtures were incubated with TKS for a short time, after which the TKS was removed by magnetic aggregation and the supernatant was transferred to a new reaction vessel containing either OAC or water. The accumulation of OA was then analyzed by LCMS as described in Section 3.3.1. A flowchart for this experiment is illustrated in Figure 5.1.

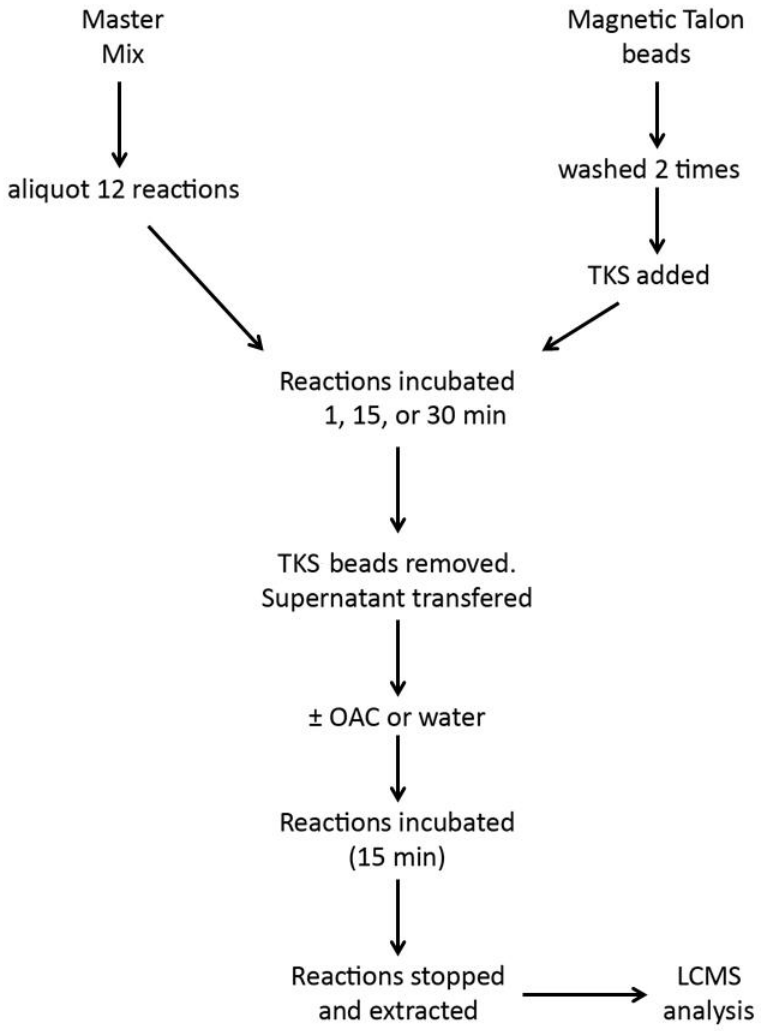


Figure 5.1 Flowchart for TKS pulldown assays.

5.2.6.1 Magnetic Talon Beads

The binding capacity is 705 µg of protein for each 1 mL suspension of Talon beads, therefore 25 µL of beads per 10 µg of TKS were used, although 25 µL of beads is capable of binding almost 2x this amount of protein (*i.e.* 18.75 µg of protein). A total of 325 µL Talon suspension (12 reactions) was pelleted (1,000 g for 1 min) and washed twice using 300 µL of 20 mM HEPES [pH 7.0] containing 5 mM DTT. A total of 25 µL of 5 mg/mL TKS was added to 275 µL of 20 mM HEPES [pH 7.0] containing 5 mM DTT. This volume was added to the washed Talon beads and was allowed to tumble for 90 min at 4°C to assure proper binding. Beads were pelleted once more (1,000 g for 1 min) and the supernatant discarded. A fresh 300 µL of buffer was added to the Talon-TKS beads. This allowed for 10 µg of TKS per 25 µL volume.

5.2.6.2 TKS Pulldown Reactions

All experiments were conducted in triplicate. Reaction mixtures (100 µL) were derived from a master mix consisting of 0.2 mM hexanoyl-CoA, 0.6 mM malonyl-CoA, 5 mM DTT, and 20 mM HEPES [pH 7.0]. Reactions were started by adding 25 µL of TKS-beads. Samples included:

Sample 1 – TKS-beads alone. Incubated at 10°C for 30 min. This sample was a negative control.

Sample 2 – TKS-beads + OAC. Incubated at 10°C for 30 min. This was a positive control.

Sample 3 – TKS-beads alone at 10°C for 15 min. Beads were pulled down, the supernatant was transferred to a new reaction vessel containing OAC and allowed to react for an additional 15 min at 10°C. This sample is the experimental sample.

Sample 4 – TKS beads alone incubated at 10°C for 1 min. Beads were removed and supernatant was transferred to a new tube, then maintained at 10°C for 14 min. OAC was added to the

supernatant and allowed to react for an additional 15 min at 10°C. This sample was to assure no TKS was being released; it represents another negative control.

Reactions were extracted twice using 500 µL ethyl acetate. Extracts were dried under vacuum then resuspended in 20 µL of resuspension solution. Injections of 10 µL were analyzed by LCMS as previously described (Section 3.1.1).

5.2.7 Equilibrium Enzyme Assays

Assays were performed using Fast Micro-Equilibrium Dialyzers with 100 µL chambers (Harvard Apparatus) (Figure 5.2). Each apparatus contained two chambers separated by a 5,000 Da MWCO cellulose acetate membrane that had been rinsed in $_{d}$H20, and then equilibrated for 30 min in 20 mM HEPES [pH 7.0]. Each chamber contained 100 µL reaction mixture consisting of 20 mM HEPES [pH 7.0], 5 mM DTT, 200 µM hexanoyl-CoA, and 600 µM malonyl-CoA. Reactions were initiated by adding TKS (20 µg) into one of the chambers. Reactions included a negative TKS-only control, a positive TKS and OAC (10 µg) control containing both enzymes in one chamber, and a TKS and OAC sample containing 20 µg TKS and 10 µg OAC in separate chambers. After incubating the reactions for 2 h at 10°C, reactions were transferred to 1.5 mL tubes and extracted using 2 x 500 µL of ethyl acetate. Extracts were dried under vacuum and resuspended in 20 µL of resuspension solution. 10 µL aliquots were analyzed by LCMS as described in Section 3.1.1.

Figure 5.2 The Harvard Apparatus Dialyzer

This dialyzer, which is similar to the one used during my experiments, separates two 100 µL chambers with a cellulose acetate membrane allowing for equilibrium to be reached between both chambers, while partitioning molecules > 5 kDa (*i.e.* TKS and OAC).

5.3 Results

5.3.1 Sub-Cellular Localization of TKS and OAC

To test whether both TKS and OAC co-localize *in planta*, fluorescent fusion constructs were made and used to infiltrate *N. benthamiana* plants where their location could be monitored by confocal microscopy. Specifically, TKS was subcloned without a stop codon into the Gateway-compatible pEarley102 vector to generate a pTKS:CFP, and OAC was subcloned without a stop codon into the pEarley101 vector, resulting in the generation of pOAC:YFP (Earley et al., 2006). These two constructs were separately used to transform *Agrobacterium tumefaciens*, which subsequently were used to infiltrate *N. benthamiana* plants based on previously reported methods (Sparkes et al., 2006; Weigel and Glazebrook, 2006). After a two-day incubation period, epidermal cells were observed using confocal microscopy, where TKS:CFP and OAC:YFP were visualized following excitation with a diode laser at the appropriate wavelengths. Results indicate that both constructs co-localized to the same sub-cellular compartment, with some residual fluorescence detected in the nucleus (Figure 5.3).

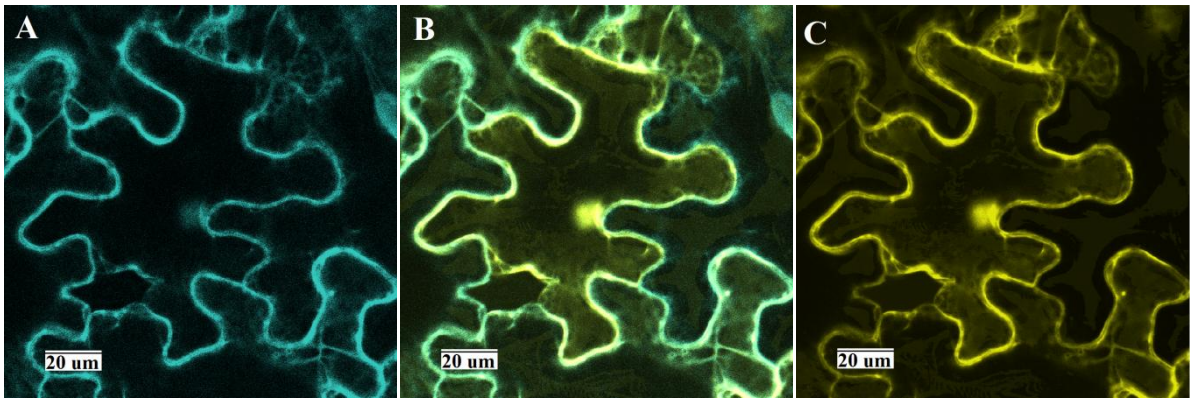


Figure 5.3 Co-localization of TKS and OAC.

The subcellular localization of (A) TKS:CFP and (C) OAC:YFP in *N. benthamiana* cells show both constructs co-localize as seen when images are (B) overlayed. *N. benthamiana* epidermal cells were examined by confocal microscopy.

5.3.2 Yeast Two-Hybrid Approach

Once the subcellular co-localization of TKS and OAC was confirmed, Y2H experiments were carried out to determine if TKS and OAC physically interact. TKS and OAC were cloned into the pDest32 and pDest22 vectors. A combination of plasmids were then used to transform the *Saccharomyces cerevisiae* strain MaV203, resulting in a collection of different yeast transformants (Table 5.5). Before proceeding with the full characterization of transformants, it was required to assess the amounts of 3-aminotriazole (3-AT) needed to inhibit HIS3 self-activation (Ausubel et al., 1994). Yeast transformants SG100, SG101, SG102, SG200, SG201, and SG202 with an OD_{600} equal to 0.1 ± 0.01 were diluted to concentrations of 0.5, 0.25, 0.125, 0.0625, and 0.03125 (based on original OD_{600}), before being spotted onto several SC-Leu-Trp-His plates containing 10 mM, 25 mM, 50 mM or 100 mM 3-AT. Plates were incubated at 30°C for three days (Figure 5.4). A concentration of 50 mM 3-AT was found sufficient to inhibit HIS3 self-activation.

Once the appropriate level of 3-AT was determined, MaV203 yeast transformants were spotted onto SC-Leu-Trp-His plates containing 50 mM 3-AT, and then grown for three days at 30°C. Other than the positive controls, no other transformant was capable of growth (Figure 5.5). Additionally, the β-galactosidase activity of transformants was also assayed, yet only the yeast in positive controls were capable of cleaving X-gal, resulting in a characteristic blue color (Figure 5.5). Collectively, these Y2H results indicate that OAC and TKS do not physically interact at the molecular level.

Table 5.5 Yeast transformants used in Y2H experiments

NAME	BAIT VECTOR	PREY VECTOR	FUNCTION
SG100	pEXP32/Krev1	pEXP22/RalGDS-wt	strong positive control
SG101	pEXP32/Krev1	pEXP22/RalGDS-m1	intermediate positive control
SG102	pEXP32/Krev1	pEXP22/RalGDS-m2	negative control
SG103	pDEST32	pDEST22	negative control
SG104	pDEST32	pSG057 (TKS)	negative control
SG105	pDEST32	pSG059 (OAC)	negative control
SG106	pSG058 (TKS)	pDEST22	negative control
SG107	pSG060 (OAC)	pDEST22	negative control
SG200	pSG058 (TKS)	pSG057 (TKS)	negative control
SG201	pSG058 (TKS)	pSG059 (OAC)	test sample
SG202	pSG060 (OAC)	pSG057 (TKS)	test sample
SG203	pSG060 (OAC)	pSG059 (OAC)	negative control

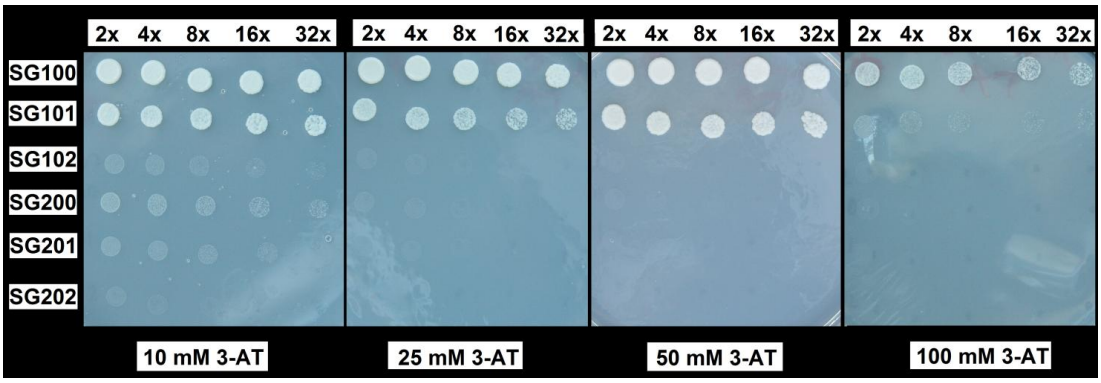


Figure 5.4 Determining 3-AT concentrations used in Y2H assays.

MaV203 yeast transformants SG100, SG101, SG102, SG200, SG201, and SG202 were spotted onto SC-Leu-Trp-His plates containing varying concentrations of 3-amino-1,2,4-triazole (3-AT).

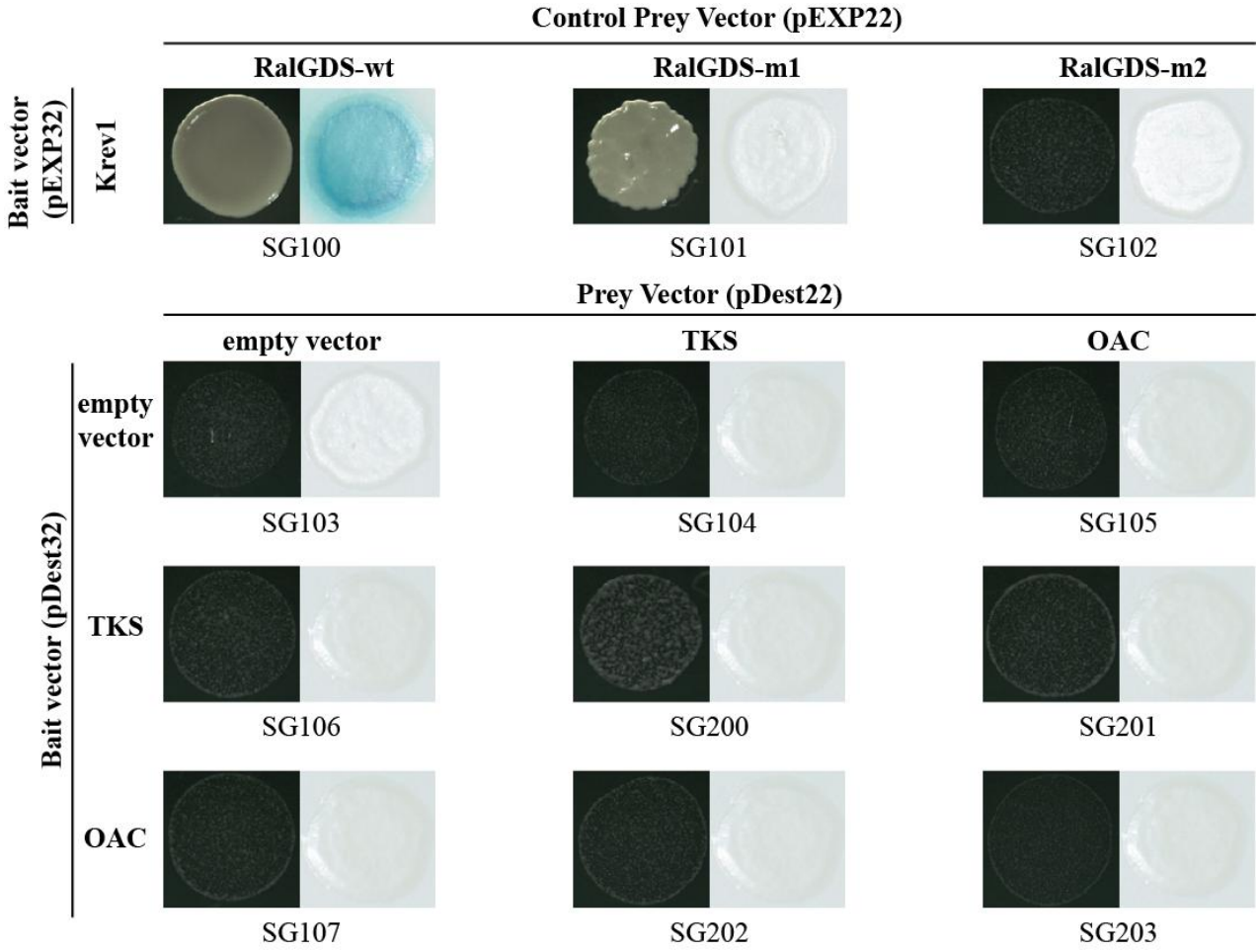


Figure 5.5 Histidine biosynthesis and β-Galactosidase activity in Y2H screens.

Yeast two-hybrid assays included strong (SG100) and intermediate (SG101) positive controls, and a negative (SG102) control. Colonies spotted on Histidine drop-out medium are shown in left panels, whereas colonies transferred to assess β-galactosidase activity are shown on the right. Numbers below the panels represent the MaV203 yeast transformant tested. Note that SG101 did appear faint blue, but the color is not clelarly visible in this photo.

5.3.3 The Expression and Purification of OAC in pJE60174

Due to the low expression of OAC in pHIS8 and pET100 vectors, OAC was codon-optimized for expression in *E. coli* using the pJexpress 411 vector (DNA2.0), resulting in the pJE60174 construct. Following the transformation of Rosetta2 BL21(DE3) Competent Cells (Novagen) and the subsequent growth of a starter culture (100 mL LB + 50 µg/mL of kanamycin), 1 L cultures

were grown for 16 h in autoinduction medium containing 50 µg/mL of kanamycin at either 25 or 37°C. Cultures were started using 50 mL starter cultures, from which a pre-induction sample for SDS-PAGE was taken (Figure 5.6). The variation in temperature during culturing was performed to determine the optimal temperature for the expression of OAC. Post-induction samples were also collected prior to purification for SDS-PAGE analysis (Figure 5.6). OAC was purified by IMAC using Talon as previously described (Section 3.1.1), resulting in ca. 1 mL protein solutions. A 5 µL aliquot of purified OAC from both 25°C and 37°C was used during SDS-PAGE (Figure 5.6). Although the Talon purification of OAC used excessive amounts of Talon to accommodate for excess protein that may or may not have been present, SDS-PAGE results indicate that the expression of OAC was increased when 25°C temperature was used instead of 37°C (Figure 5.6). Expression of OAC at lower tempertatures resulted in ca. 8.5 mg of protein, whereas expression at 37°C yielded ca. 1.2 mg of OAC. Thrombin was also used to remove the poly-histidine tag from OAC (Figure 5.7).

One evident benefit of the inclusion of OAC in the pJexpress 411 vector was the loss of the dual band seen in previous OAC purifications derived from pHIS8 and pET100 vectors (Figure 4.1). The purification of OAC from pJexpress-based vectors resulted in the appearance of a single band during SDS-PAGE (Figure 5.7)

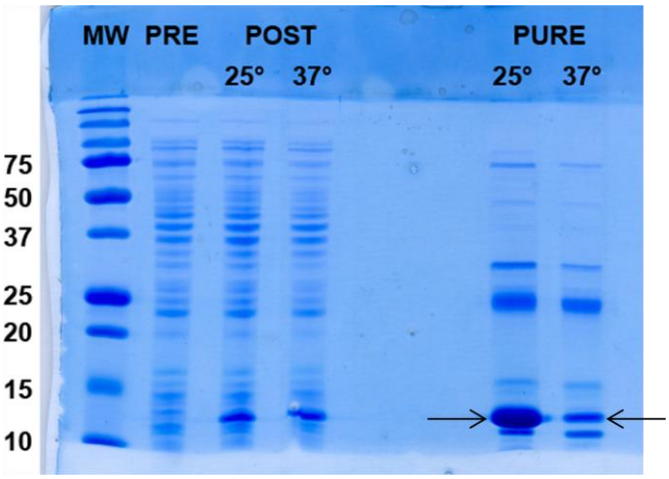


Figure 5.6 SDS-PAGE of OAC expressed from the pJexpress 411 vector.

A 15% polyacrylamide gel was used to resolve 5.0 µL of pre-induced (PRE) and post-induced (POST) cultures expressing OAC, and Talon-purified OAC (PURE). Cultures were initiated using the same starter culture, but were then grown in parallel at either 25 or 37°C. Equal volumes were used to load the gel. Proteins were purified in the presence of excess Talon (Clontech) to ensure maximal amounts of protein were retained. OAC is evidently expressed, as is seen in the PURE 25°C lane at the 13 kDa mark. OAC is indicated by arrows. Numbers on the left of the gel correspond to the kDa of molecular weight markers (MW).

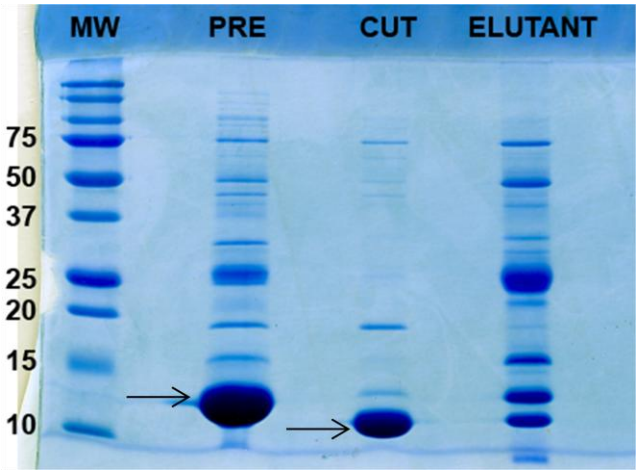


Figure 5.7 SDS-PAGE of Thrombin cut OAC.

IMAC purified OAC (PRE) was digested in the presence of Thrombin resulting in the loss of its polyhistidine tag. Digested OAC (CUT) was isolated by running the solution over Talon beads, which adhered to uncut OAC and other contaminants. The Talon-bound contaminants were also eluted and visualized on this gel (ELUTANT). The gel was overloaded in order to detect contaminating protein. Equal volumes of solution were used to load this gel. OAC is indicated by arrows. Numbers on the left correspond to the kDa of molecular weight markers (MW).

5.3.4 TKS Magnetic Pulldowns

Experiments were designed to address whether or not OAC is capable of cyclizing TKS products in the absence of TKS. The experiment allowed TKS to catalyze polyketide synthesis for a brief time before being magnetically removed (pulled down), and removing the supernatant containing TKS products. The supernatant was then incubated with OAC or water, extracted, and analyzed by LCMS. Reactions included: 1) a negative control using only TKS-beads (B-TKS) that should be capable of producing HTAL, PDAL, and olivetol, but not OA; 2) a positive control that assayed TKS-beads in the presence of OAC and should be capable of producing OA; 3) an experimental sample that allowed TKS-beads to react for 15 min before using a magnet to pull the TKS-beads out of solution, then transferring the supernatant to a new reaction vessel containing OAC and incubating the solution for another 15 min, and 4) a TKS bead leak control that pulled down the TKS beads prior to transferring the solution to a reaction vessel containing malonyl-CoA, then allowing the reaction to proceed for another 15 min. These assays were also performed in parallel with assays that used soluble TKS (S-TKS) to compare the differences in product accumulation between assays. Initially, to determine the optimum temperature at which to conduct these experiments, the assays were performed at 10, 20, and 30°C (Figure 5.8). Results show that TKS-only negative controls only produced HTAL, PDAL, and olivetol, whereas TKS and OAC positive controls produced HTAL, PDAL, olivetol, and OA. A difference between S-TKS and B-TKS chemoselectivity was observed, where S-TKS produced almost an order of magnitude more HTAL and PDAL than did B-TKS, in the absence or presence of OAC and at all temperatures (Figures 5.8 and 5.9). Notably, the amounts of olivetol

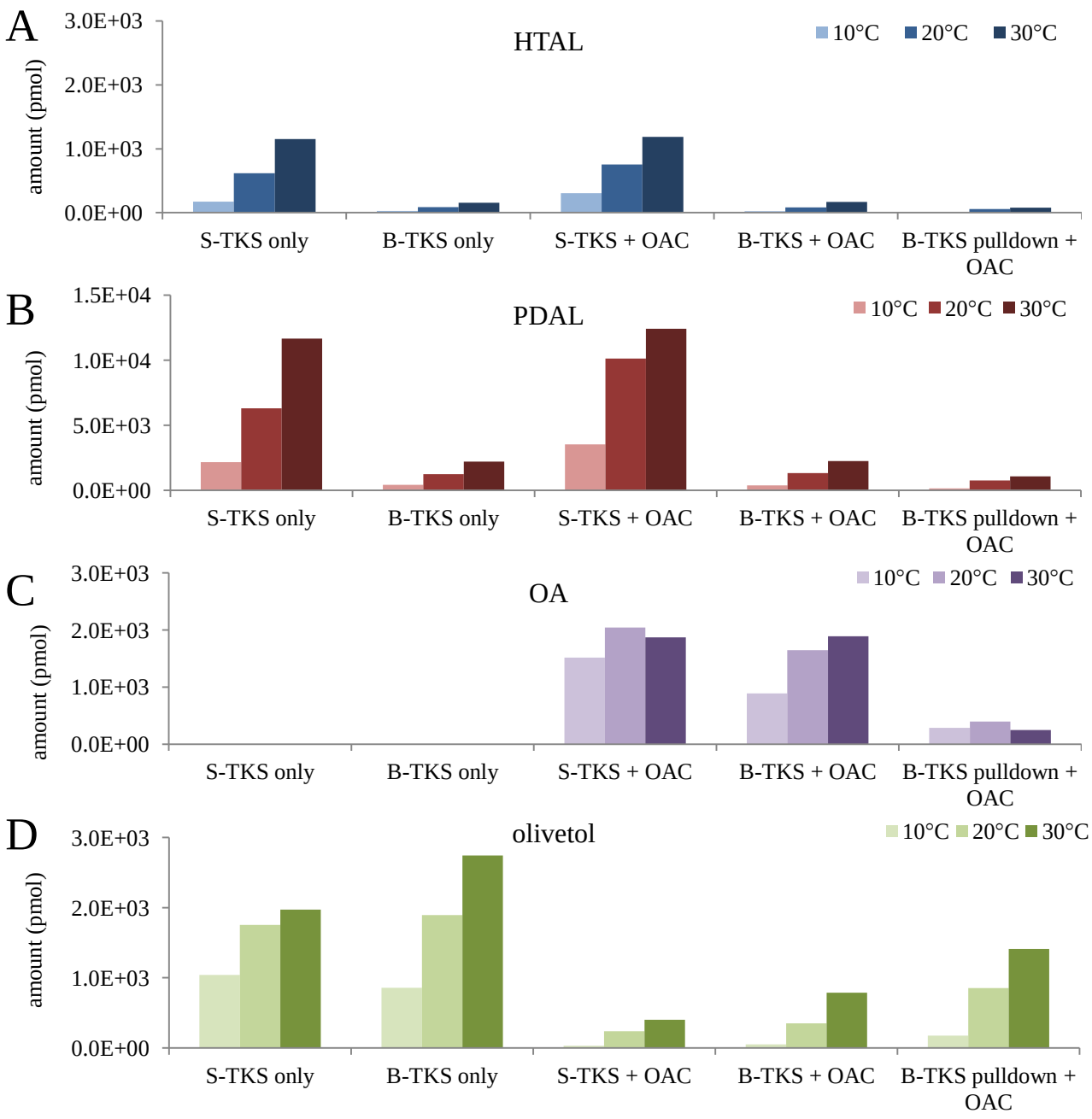


Figure 5.8 Products accumulating during TKS pulldown assays.

The product yields from soluble (S-TKS) and Talon-bound (B-TKS) TKS assays performed at 10, 20, and 30°C in the presence or absence of OAC. Each panel corresponds to one reaction product, where the products were (A) HTAL, (B) PDAL, (C) OA, and (D) olivetol.

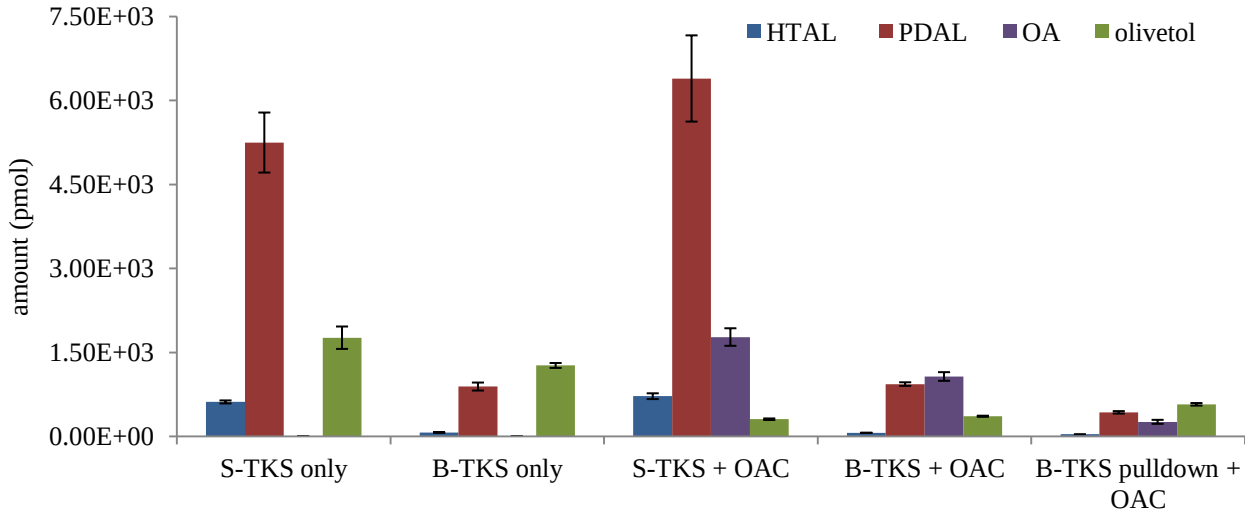


Figure 5.9 The production of OA by OAC following TKS pulldown.

A comparison of soluble (S-TKS) and Talon-bound (B-TKS) TKS assays conducted in the presence and absence of OAC show product amounts obtained when B-TKS was pulled out of solution before adding OAC. All assays that included OAC successfully produced OA. Error bars are the SD where n=3.

produced by S-TKS and B-TKS were approximately the same, which accumulated to ca. 1,800 pmol when assayed in the absence of OAC, and to ca. 300 pmol when assayed in the presence of OAC (Figures 5.8 and 5.9). Similarly, OA production was also approximately the same when assaying S-TKS or B-TKS in the presence of OAC, where the amount of OA accumulated to ca. 1800 pmol (Figures 5.8 and 5.9). Samples where B-TKS was removed prior to the addition of OAC did accumulate OA. These samples accumulated OA ca. four times less than B-TKS + OAC positive controls, accompanied by ca. 2.5 times more olivetol than seen in positive controls. Optimum temperature for OA production differed between S-TKS and B-TKS samples, where S-TKS showed increased OA activity at 20°C, and B-TKS showed optimum OA activity at 30°C. Samples that were designated as B-TKS leak controls did not show any activity. Once an optimal temperature of 20°C was determined (Figure 5.8), experiments were repeated to confirm results (Figure 5.9).

The hypothesis that TKS produces a soluble linear tetraketide intermediate was also explored by monitoring the TKS-B reactions over time (Figure 5.10). Reactions were incubated with TKS-B for 15 min before pulling down TKS and transferring the supernatant to a new reaction vessel. OAC was added to reactions immediately, or after 5, 10, 20 or 40 min had elapsed. OAC reactions were incubated 15 min before stopping and extracting the reactions. These reactions also used a negative control consisting of B-TKS only, and a positive control which assayed B-TKS in the presence of OAC; both assays were incubated for 30 min before stopping and extracting them, and were conducted in parallel with the other timed assays described. All assays were conducted in triplicate. The largest amount of OA (1,600 pmol) was detected in TKS and OAC positive controls, whereas no OA was detected in TKS-only negative controls. Pulldown assays that added OAC immediately following the removal of TKS produced the largest amount of OA (490 pmol), followed decreasingly by the 5, 10, 20, and 40 min timed samples, which accumulated 287, 256, 238 and 200 pmol of OA, respectively. OAC activity, or more correctly the substrate for OAC, was retained even after 40 min had elapsed between the time TKS was removed and the time OAC was added. Conversely, amongst the samples that were allowed to stand for a given time before OAC was added, the largest amount of olivetol (1,300 pmol) was detected in samples that stood for 40 min before the addition of OAC, and decreased as the sitting time before the addiction of OAC was shortened. Whereas HTAL amounts remained relatively static (41 to 56 pmol), the largest amount of PDAL (754 pmol) was detected in positive controls that assayed TKS in the presence of OAC. The accumulation of PDAL appeared to increase as sampling time increased, although the initial 0 min sample did contain an intermediate amount of PDAL (*i.e.* more than the 5 and 10 min samples, but less than

the 20 and 40 min samples). The least amount of PDAL and HTAL were detected in negative controls.

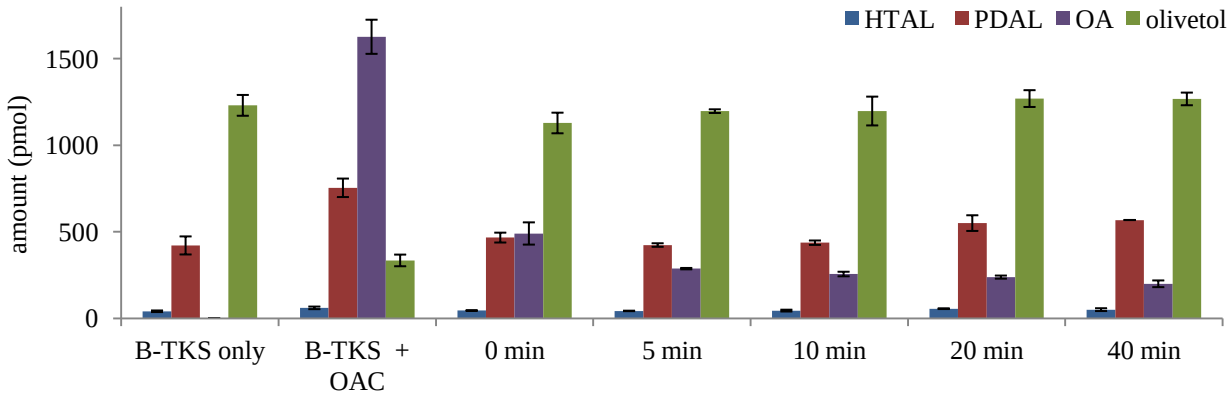


Figure 5.10 Monitoring OAC substrate availability following TKS pulldowns.

Talon bound TKS (B-TKS) was incubated for 15 min in a reaction mixture before being removed by magnetic pull-down. Following the transfer of the supernatant, OAC was added after a given time (0 to 40 min), and allowed to react for another 15 min. Error bars are the SD where n = 3.

5.3.5 The TKS/OAC Dialysis Reaction

To assess whether or not TKS and OAC can function independently, and if OAC's substrate is a diffusible intermediate that is initially produced and released by TKS before being cyclized by OAC, dialysis experiments separated TKS and OAC enzymes using a 5 kDa membrane which allowed the diffusion of substrates and reaction products between chambers. A negative control incapable of OA production was accomplished by only adding TKS to one chamber, where the second chamber contained no enzyme. A positive control was produced by adding both TKS and OAC to one chamber, and not adding any enzyme to the second chamber. Experimental samples included TKS in one chamber and OAC in the other chamber. Following 1 h incubation, each chamber's solution was separately extracted, and analyzed by LCMS. Whereas TKS-only

controls produced only HTAL, PDAL, and olivetol, positive controls and test samples were capable of producing OA in addition to the other TKS products (Figure 5.11).

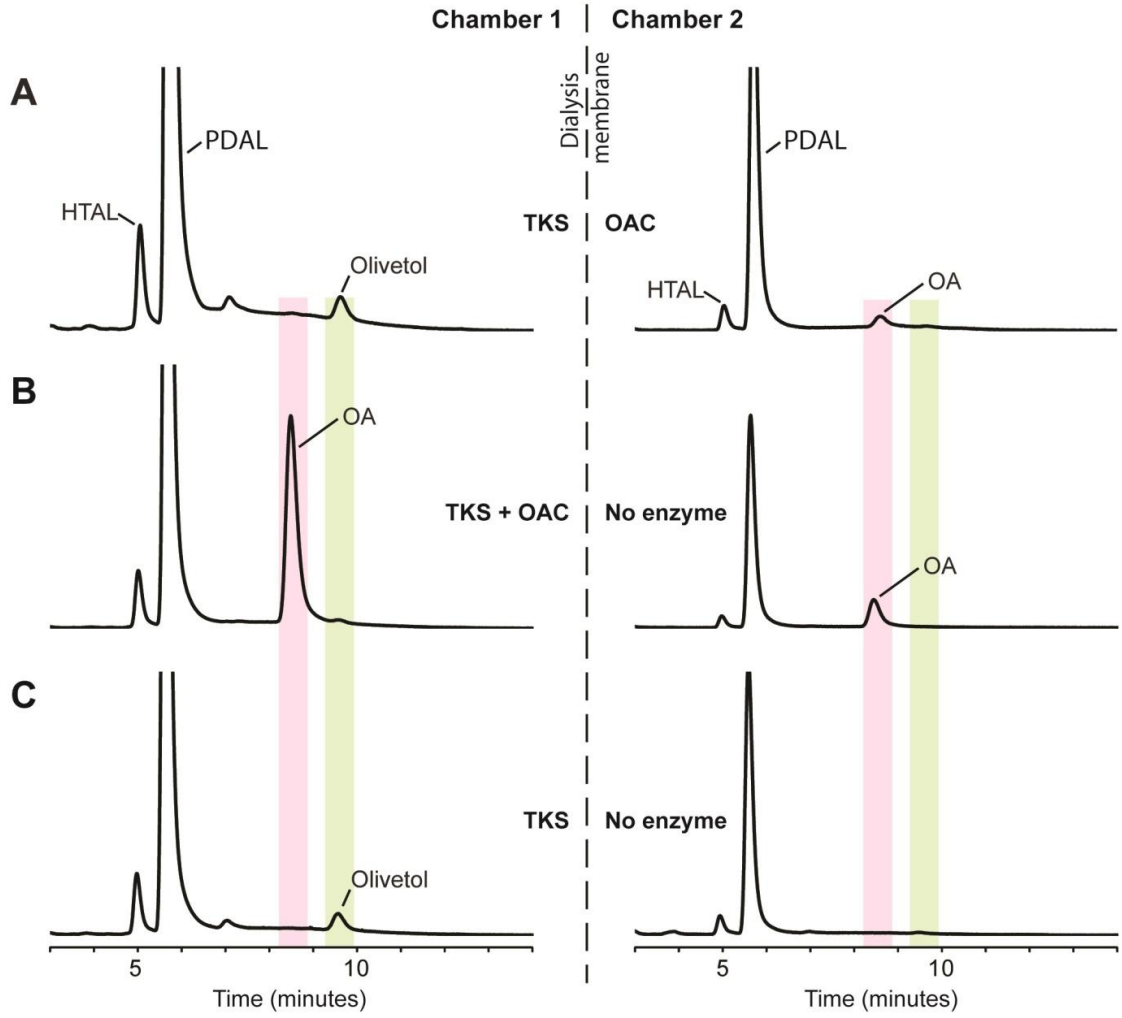


Figure 5.11 The micro-dialysis of TKS assay products.

Chromatograms are results from HPLC sample analysis that were monitored at 270 nm. (A) The use of a micro-dialysis apparatus allowed TKS/OAC assays to be conducted while enzymes were physically separated by a 5 kDa membrane in two separate chambers. (B) Positive controls included both enzymes in one chamber and resulted in the production of HTAL, PDAL, OA, and olivetol. (C) Negative controls only used TKS in one chamber. The elution of OA is highlighted in pink across samples, whereas the elution of olivetol is highlighted in green.

5.4 Discussion

A central focus of this work was to determine if TKS and OAC interact. Results from subcellular localization experiments indicate that both TKS and OAC co-localize *in planta,* and are not spatially constrained, indicating that both proteins could potentially participate in the *in vivo* production of OA (Figure 5.3). This observation raised the question as to whether or not TKS and OAC physically interact, conceivably modifying TKS activity by an allosteric mechanism, or possibly by directing a transfer of an unstable reaction intermediate. Y2H assays did not indicate that any physical interaction exists between TKS and OAC (Figure 5.5). The lack of physical interaction between these enzymes was also observed during i) TKS pulldown experiments, and ii) TKS/OAC dialysis assays, both of which physically separated TKS and OAC during assays, yet were capable of producing OA. Although these latter experiments cannot directly determine if TKS and OAC physically interact, they show that TKS products can be cyclized by OAC in the absence of TKS, which shows that these enzymes do not require physical interaction for OA production. These results also indicate that neither an allosteric mechanism nor the direct transfer of polyketide intermediates is required for OA biosynthesis. Instead, these results, along with the observation that neither PDAL, HTAL, nor olivetol are appropriate substrates for OAC (Section 4.2.10.2), and that OA production is augmented at colder temperatures (Section 4.2.5) suggest that a somewhat thermolabile tetraketide intermediate is released by TKS and is stable long enough for OAC to cyclize it. These findings challenge the current understanding of type III polyketide cyclization, which was understood to occur exclusively within the PKS active site itself (Austin and Noel, 2003). Namely, the Claisen condensation that occurs in CHS has been described as an intramolecular cyclization that results

in the release of the aromatic compound from either a conserved cysteine that tethers the nascent polyketide chain to the enzyme, or from CoA. Although the mechanism of polyketide release from the enzyme is currently unknown, the cyclization specificity of chalcone and stilbene synthases has been attributed to specific amino acid residues found within the type III PKS active site (Austin et al., 2004). There exists evidence in support of a short-lived polyketide intermediate, especially in type II PKSs, which produce irregular shunt products when they are assayed in the absence of the necessary polyketide cyclases (Shen and Hutchinson, 1996; Sultana et al., 2004). Since TKS does not cluster with traditional CHS or STS, and is instead related to unique non-chalcone synthase PKSs, it is likely that the mechanism of retention, release and cyclization of polyketide intermediates by TKS differs from traditional CHS/STS paradigms (a descriptive analysis of the phylogenetic relationship between PKSs is shown in Appendix G (Schröder, 2010).

The pathway for OA biosynthesis in cannabis requires the involvement of OAC, a unique polyketide cyclase, which has not been previously known to occur in plants. Instead, similar cyclases have been found in bacterial type II PKS systems, where multiple enzymes participate in the biosynthesis of polyketide products (Shen and Hutchinson, 1996). The involvement of OAC-like cyclases during polyketide cyclization may not be exclusive to cannabis; it is conceivable that similar cyclases are involved in other plants where the accumulation of carboxylated polyketides remains unclarified. Examples of resorcylic acid accumulation in plants include anacardic acids in cashew and stilbene carboxylates in *Hydrangea* (Gorham, 1995; Kozubek and Tyman, 1999), although it should be noted that type III PKSs capable of producing

long-chained alkylresorcylic acids without the requirement of a cyclase have been previously identified in rice and in the mold *Neurospora crassa* (Funa et al., 2007; Matsuzawa et al., 2010).

An observation indirectly related to this project was the reduction of lactone products during assays performed using Talon-bound TKS. Assays that compared the activities of immobilized and soluble TKS showed a decrease in the production of lactone products (HTAL and PDAL) with almost no change to OA production when immobilized TKS was used (Section 5.2.3). Similar results have also been reported for alfalfa CHS that had been immobilized onto Ni-NTA agarose, which, in addition to increased tetraketide:triketide yields, resulted in a bound enzyme with a much more effective half-life than soluble CHS (Kim et al., 2009). Although the half-life of TKS was never explored in the current study, the observation that both immobilized PKSs produce higher tetraketide:triketide yields suggests that product derailment occurs easily in these enzymes, and that *in vivo* these enzymes may very well be anchored to a naturally-occurring support matrix, as suggested for phenylpropanoid and flavonoid metabolism (Winkel-Shirley, 1999). Recent studies have shown that resuspended type II minimal PKS involved in actinorhodin biosynthesis increases product yield when retained and stabilized within a dense coacervate matrix (Crosby et al., 2012). The authors attribute their findings to a crowding effect that allows for the direct transfer of PKS intermediates, but it may also be due to the stabilization of the PKS itself which leads to its increased efficiency. The matrix used by Crosby et al. (2012) is also reminiscent of the unique physiology of the trichome cell, which is dominated by spherical secretory bodies that could potentially be involved in anchoring TKS (Hammond and Mahlberg, 1978). The chemoselective adaptability observed in TKS could be of practical benefit in other type III PKS research that hopes to maximize enzyme efficiency.

6. THE STRUCTURE AND MECHANISM OF OAC

6.1 Introduction

In the previous chapter, experiments were conducted to examine the physical relationship shared between TKS and OAC. The observation that TKS and OAC contructs co-localize *in vivo* in *N. benthamiana* suggested that the proteins could potentially interact, which led to yeast two-hybrid experiments being performed to determine if physical interaction did occur. Yeast two-hybrid assays indicated that no physical interaction exists between TKS and OAC. In turn, this observation led to experiments that separated TKS and OAC, including the physical removal of TKS midway through a reaction by magnetic pull-down methods, as well as the use of a micro dialysis apparatus that allowed the enzymes to be separated by a permeable membrane. These experiments indicated that the enzymes perform separate reactions, wherein TKS synthesizes a theoretical soluble tetraketide intermediate, which is then independently cyclized to OA by OAC.

The observations that TKS and OAC act separately has led to two related and important questions, which are; what is the mechanism of cyclization used by OAC, and what is the nature of the theoretical tetraketide intermediate? In the present chapter, experiments that were conducted to address the mechanism of OA cyclization and the nature of the intermediate are presented.

The mechanism of chalcone and stilbene biosynthesis involves an intramolecular cyclization step that can be dramatically influenced by replacing several key amino acids located in the active site (Austin and Noel, 2003). Comparatively, TKS appears to behave as an STS when assayed alone. Similarly, during TKS and STS cyclizations, a carboxyl moiety is lost,

which suggests that an aldol reaction occurs during the production of olivetol as it does during the biosynthesis of stilbenes. This is contrasted by Claisen reactions that occur in CHSs, which retain all carbon moieties during the cyclization of final products. Yet, when TKS is assayed in the presence of OAC, the two enzymes together behave as an STCS capable of catalyzing an aldol C2 → C7 cyclization while retaining its C_1 carboxyl moiety (Figure 2.5). To better understand the mechanism of cyclization adopted by OAC, traditional aldolase mechanisms, which include class I and class II reactions, were considered (Gijsen et al., 1996). The role of a metal as a catalyst could have implicated a class II aldol mechanism in OAC, but this was dismissed since no metal is required for OA cyclization (Chapter 4), suggesting that a type I aldolase mechanism could be employed by OAC. It was a goal of this research to examine the lysine residues putatively involved in a type I aldolase mechanism in OAC, which were conserved by homologs that could share a similar function, but in addition to simply investigating conserved lysine residues, several AA residues that are conserved in OAC and its homologs were also investigated.

The experiments in the present chapter use the primary and secondary structure of OAC to determine residues putatively involved with enzyme activity. The primary sequence of OAC was compared with twelve other SRABB proteins, and a homology-based model of OAC was developed. These analyses led to the identification of conserved AAs shared between SRABB homologs, some of which were subsequently targeted for alanine substitution. Additionally, to better understand the putative active site, the OAC structure was modeled using the crystal structures of *A. thaliana* HS1 (pdb 1Q53) and *P. tremula* SP1 (pdb 1TR0) as templates, allowing

the closer inspection of the enzyme's three-dimensional strucutre and the positions of site-directed mutatgenesis (SDM) sites.

Since previous observations regarding TKS and OAC reactions imply the existence of a transient tetraketide intermediate, attempts to detect its existence were performed using highly sensitive analytical MS techniques. MS ion screens were performed during TKS and TKS/OAC reactions in order to detect CoA-linked compounds, with specific interest in a tetraketide-CoA intermediate that could exist transiently in solution. Due to the uncertainty about the reaction mechanism and in determining which intermediate is produced during a TKS reaction, these techniques were used to find ions possessing CoA fragments with a common positively charged fragment of 428 m/z (Figure 6.1). Similarly, neutral loss experiments were also used to detect a loss of 507 amu from CoA-containing molecules.

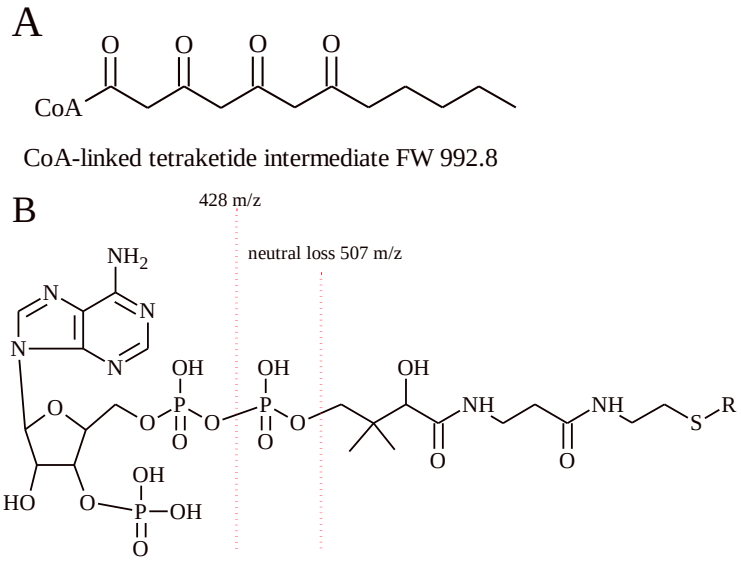


Figure 6.1 CoA fragmentation during MS analysis.

(A) A proposed linear polyketide intermediate with a molecular weight of 992.8 screened for during MS analysis. (B) Sites of bond cleavage occur along the dashed lines in the CoA-portion of acyl-CoAs during MS analysis conducted in positive ion mode. Substituents (R) may include hexanoyl, acetyl, malonyl, and other products, such as seen in (A).

6.2 Materials and Methods

6.2.1 Primary and Secondary Structure Analysis

A BLASTP query using the OAC ORF was conducted against the NCBI non-redundant protein sequence database using a BLOSUM62 scoring matrix, a gap cost of 11 and extension cost of 1, using the conditional compositional score matrix adjustment (Altschul et al., 1997). Highest-ranking homologs with the lowest E-value were selected, although only one homolog from a species was selected. Sequences were aligned using ClustalW engine as previously described (Section 4.2.1). Analysis was conducted by querying the OAC sequence using the Swiss-Model Template Identifier (Arnold et al., 2006). The Interpro Domain Scan was conducted against the HMMPfam, HMMTigr, ProfileScan, SuperFamily, and BlastProDom databases using a gapped blast query with an E-value cutoff of 1.0E-6 and a BLOSUM62 matrix (Zdobnov and Apweiler, 2001). Secondary Structure prediction and domain assignment were completed using the PSIPRED and DISOPRED engines provided by the domain annotation feature in the Swiss-Model Workspace (Jones, 1999). The DISOPRED E-value cutoff was 2%.

6.2.2 Cloning, Expression and Activity of OAC Mutants

This work used an OAC construct (pJ60174) that was optimized for codon usage in *E. coli* and was commercially prepared in the pJE411 vector (DNA2.0). This OAC construct also contains an N-terminal polyhistidine tag for IMAC purification. In addition to the wild-type OAC (wtOAC), 12 constructs each containing a single point alanine substitution were also ordered (DNA2.0). Additional single point OAC mutants at K4A, K38A, and D45A were not commercially produced, but created by the site-directed mutagenesis (SDM) of pJ60174, and are

found in pJ60187, pJ60188, and pJ60189, respectively. Primers used during SDM are shown in Table 6.1.

Table 6.1 Primers used for SDM and sequencing of pJR60174

Name	Sequence	Function
F-K4A	GTAGCATGGCGGTGGCACACTTGATCGTG	(F) primer for pJ60187 K4A
R-K4A	CACGATCAAGTGTGCCACCGCCATGCTAC	(R) primer for pJ60187 K4A
F-K38A	CATTATTCCGGCGATGGCAGACGTTTATTGGGG	(F) primer for pJ60188 K37A
R-K38A	CCCCAATAAACGTCTGCCATCGCCGGAATAATG	(R) primer for pJ60188 K37A
F-D45A	GTTTATTGGGGCAAGGCTGTGACCCAGAAGAAC	(F) primer for pJ60189 D44A
R-D45A	GTTCTTCTGGGTCACAGCCTTGCCCCAATAAAC	(R) primer for pJ60189 D44A
F-pJE	CTCGAAAATAATAAAGGGAAAATCAG	(F) primer for sequencing
R-pJE	TGGTAGTGTGGGGACTC	(R) primer for sequencing

6.2.3 Creation of K4A, K38A, and D45A by SDM

Reactions (50 µL) were performed using 5 µL Pfu II Ultra HS 10x buffer, 1 µL of 50 ng/µL pJE60174 (template), 1 µL of 125 ng/µL forward (F) primer, 1 µL of 125 ng/µL reverse (R) primer, 0.5 µL 100 mM dNTP (25 mM of each), 40.5 µL water, 1 µL of 2 U/µL Pfu II ultra HS polymerase (Stratagene). Forward and reverse primers are described in Table 6.1. Thermocycling was performed in five steps, where steps 2 - 4 were repeated 16 times before ending with step 5: step 1 - 95°C for 1 min; step 2 - 95°C for 1 min; step 3 - 58°C for 1 min; step 4 - 68°C for 2 min; step 5 - 68°C for 3 min. Reactions were incubated at 37°C for 60 min in the presence of 1 µL of 8 U/µL Dpn1 (Sigma).

Plasmids from SDM reactions were isolated using a MinElute Reaction Cleanup Kit (Qiagen), and each used to transform separate *E. coli* One Shot BL21 (DE3) competent cells (Invitrogen). Transformants were plated onto LB + 50 µg/mL kanamycin and cultured overnight at 37°C. Six colonies from each transformation were selected and grown in 5 mL LB + 50

mg/mL kanamycin overnight at 37°C. Plasmids were purified using a QIAprep Spin Miniprep Kit (Qiagen) and sequenced (PBI DNA Technologies Unit) using the F-pJE and R-pJE oligonucleotides (Table 6.1) to confirm that the SDM reactions were successful.

6.2.3.1 Expression and Purification of OAC and its Mutants

Plasmids harboring wild-type OAC (wtOAC) and OAC mutants were individually used to transform *E. coli* Rosetta (DE3) competent cells (Novagen). Transformation reactions were plated on LB + 50 µg/mL kanamycin and grown overnight at 37°C. Single colonies were used to inoculate two 5 mL cultures of LB + kanamycin, where one solution was used as a starter culture and the other for i) -80°C stocks (as described in Section 3.1.2), and ii) plasmid isolation and sequencing as described in Section 6.2.3. A 0.5 mL aliquot of starter culture was used to inoculate 50 mL LB + 50 µg/mL kanamycin, which was grown overnight at 37°C with shaking (200 RPM). Cultures were used to inoculate 500 mL of LB + 50 µg/mL kanamycin, and grown to an OD_{600} = 0.8 before being induced using 0.5 mM IPTG. Cells were grown for an additional 5 h at 37°C before being harvested by centrifugation at 20,000 g. Pellets were stored overnight at -80°C.

Cell pellets were thawed on ice in the presence of 50 mL lysis buffer. The solution was divided into 25 mL aliquots and each sonicated for 3 min on ice. The solutions were then centrifuged for 20 min at 20,000 g at 4°C. The decanted lysates were tumbled in the presence of 500 µL suspension volume of Talon (Clontech) for 40 min at 4°C. The resin was centrifuged (1 min at 1,000 g) and the supernatant removed. The Talon beads were resuspended in 10 mL wash buffer and tumbled at 4°C for 10 min to wash away protein impurities. The washing step was conducted four more times before the Talon beads were rinsed in a gravity flow column with a

10 mL wash buffer. Each protein was eluted using 10 mL of elution buffer directly into separate Amicon Ultra 15 centrifugal concentrators (5,000 MWCO) (Millipore), which were later used to concentrate the solutions to 2.5 mL. The solutions were desalted using a PD10 column (GE Healthcare Biosciences) equilibrated with storage buffer (20 mM HEPES pH 7.0, 25 mM NaCl, 10% glycerol, 5 mM DTT). The solutions were further concentrated to 100 µL using Microcon centrifugal filter devices (Millipore). These solutions were quantified using an RC/DC protein assay kit (Bio-Rad) as directed by the manufacturer.

6.2.3.2 Enzyme Assays

Separate 50 µL reactions containing a final concentration of 0.2 mM CoA, 0.4 mM ATP, 2.5 mM $MgCl_2$, 8 mM sodium malonate, 5 mM DTT, 20 mM HEPES [pH 7.0], and 0.2 mM hexanoyl-CoA were prepared using a master mix. Each reaction mixture contained 200 pmol of wtOAC or one of the single point OAC mutants. The first experiments (experiment A) did not include the K4A, K38A, or D45A mutants, whereas the second set of experiments (experiment B) did. Experiments A and B were conducted independently with slight differences described in the methods below. All reactions were conducted in triplicate.

Experiment A

Reactions were started by adding 2 µL of an enzyme cocktail containing 100 pmol TKS and 200 pmol MCS. Reactions were incubated for 30 min at RT before being extracted twice using 250 µL ethyl acetate. Extracts were dried under vacuum and analyzed by LCMS as previously described (Section 3.1.1).

Experiment B

Reactions were started by adding an enzyme cocktail containing 200 pmol of TKS and 200 pmol of MCS, and were incubated at 20°C for 45 min before being extracted twice using 200 µL ethyl acetate. Extracts were dried under vacuum and stored overnight at 4°C. The extracts were resuspended in 20 µL of resuspension solution. 10 µL aliquots were analyzed by LCMS as previously described (Section 3.1.1).

6.2.4 MS/MS Monitoring of the TKS/OAC Reaction

Reaction mixtures of 50 µL were prepared with 0.6 mM malonyl-CoA, 0.2 mM hexanoyl-CoA, 5 mg/mL enzyme (TKS and/or OAC) in 1 mM HEPES [pH 7.0] and 5 mM DDT, and were incubated for 1 h at 37°C, before being stored in the autosampler at 8°C. A control sample was also prepared without enzyme present. Reaction mixtures were diluted by a factor of ten with 1 mM HEPES [pH 7.0] and injected to a stream of 50:50 ACN:H_2O with 5 mM triethylamine (TEA) and 3 mM acetic acid (HOAc), using an Alliance 2695 HPLC coupled to a Quattro Ultima mass spectrometer (Waters Corporation). Signals were recorded in positive ion mode electrospray ionization with tandem MS scanning of selected reaction monitoring (SRM) for each of the acyl coenzyme-A compounds of interest; malonyl-CoA, acetyl-CoA, hexanoyl-CoA, free CoA, formyl-CoA, and a tetraketide intermediate-CoA (992 m/z) (de Hoffmann 1996). Since an intermediate was expected, an alternate mode of tandem MS precursor ion scanning was also used to monitor any possible acyl-CoA that might be produced during the reaction. Analyses were conducted by S. Ambrose (Applied Genomics and Analytical Technologies, NRC-PBI).

6.3 Results

6.3.1 Primary and Secondary Structure Analysis

Following a BLASTP query of the OAC protein sequence, 12 representative SRABB proteins were selected and aligned using ClustalW with the MEGA5 software suite (Figure 6.2). Most SRABB homologs used in this work are annotated in Genbank as POP3 proteins and include members from: *Medicago truncatula* (annotated XP_003609973.1), *Vitis vinifera* (annotated XP_002272048.1), *Glycine max* (annotated XP_003526484.1), *Zea mays* (annotated NP_001152608.1), *Ricinus communis* (annotated XP_002510283.1), *A. thaliana* (pdb 1Q53), *P. trichocarpa* (annotated ABK93400.1), *Brachypodium distachyon* (XP_003579052.1), *Oryza sativa* (annotated NP_001043194.1), *Sorghum bicolor* (annotated XP_002458107.1), *Physcomitrella patens* (annotated XP_001766578.1) and *P. tremula* (pdb 1TR0). A total of 19 of the 103 amino acids in OAC (*i.e.* 18%) are identical in all of the selected homologs, and another 11 amino acids share similarity (Figure 6.2).

Furthermore, the OAC protein sequence was analyzed using software provided by the Swiss Institute of Bioinformatics (Arnold et al., 2006). The secondary structure was deduced using InterproScan, PSIPRED and DISOPRED capable of domain recognition, secondary structure prediction, and protein disorder determination, respectively (Figure 6.3). Interpro identified both the DABB (IPR011008) and SRABB (IPR013097) domains in OAC between amino acids 3 to 99 (Zdobnov and Apweiler, 2001). PSIPRED results indicated that OAC possesses multiple α-helices and β-pleats that fit an α + β classification (Jones, 1999). Results from DISOPRED strongly suggested that OAC is an ordered protein with a stable secondary structure possessing low flexibility in solution (Ward et al., 2004).

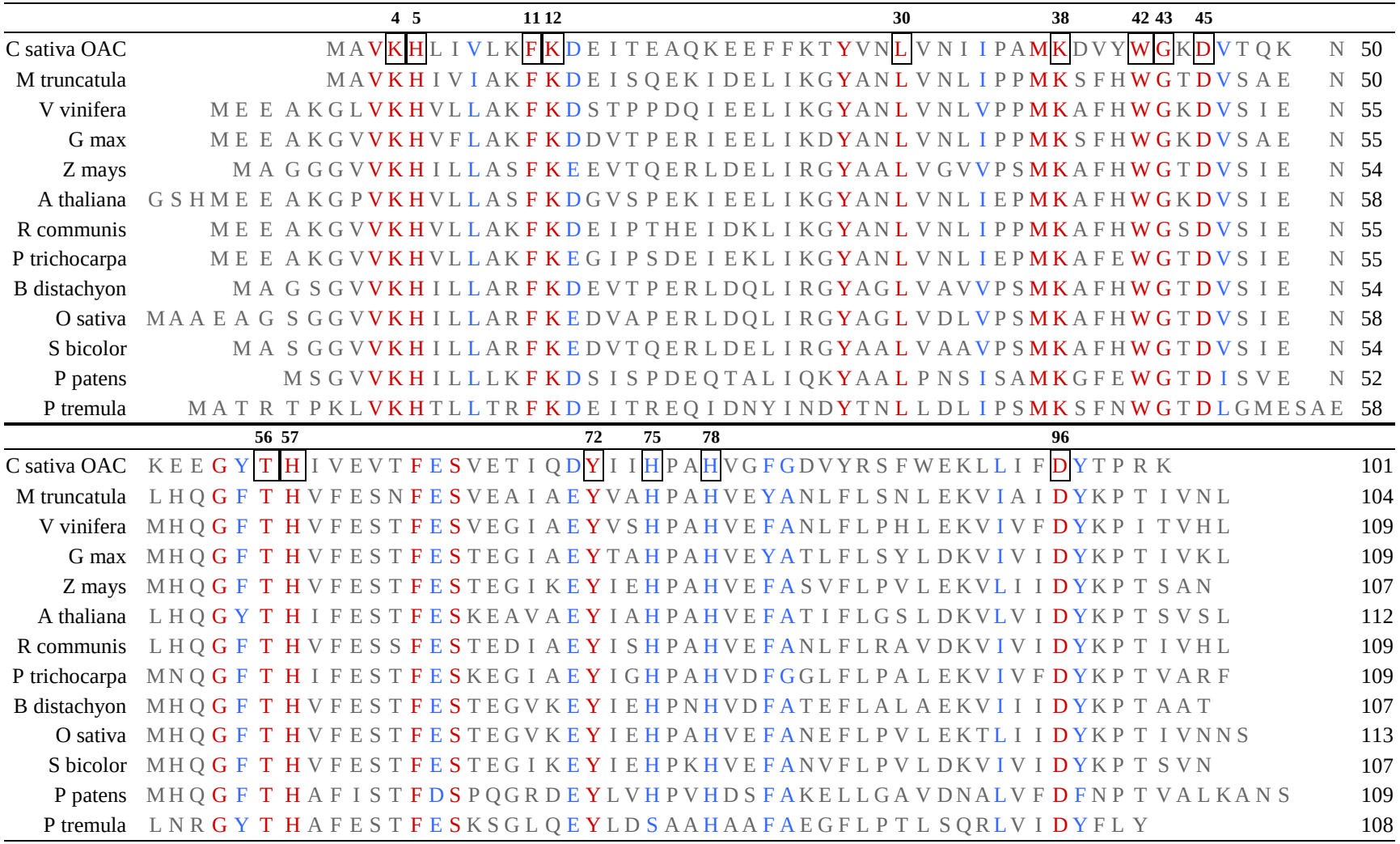


Figure 6.2 Sequence alignment of OAC and homologs.
This alignment was used to determine conserved amino acid residues, some of which (boxed) were targeted for SDM. Conserved residues that are identical are shown in red, whereas amino acids that share similarity are drawn in blue.

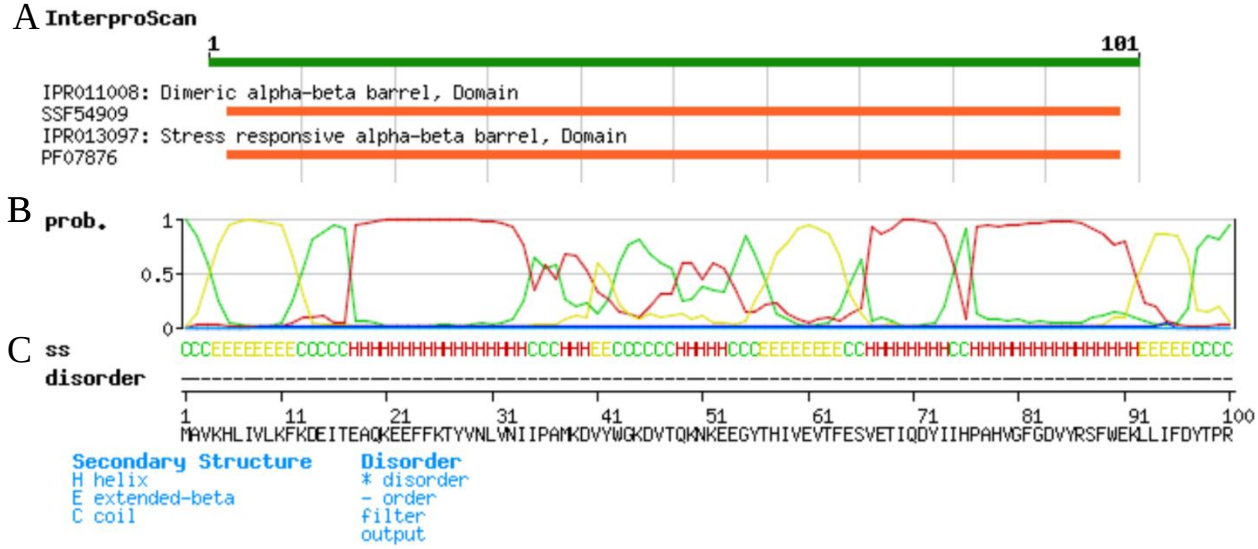


Figure 6.3 Prediction of OAC Secondary Structure.

The OAC protein sequence was evaluated using: (A) Interpro, (B) PSIPRED, and (C) DISOPRED. (A) Interpro identified the DABB and SRABB motifs in OAC. (B) PSIPRED, which is capable of predicting the probablilty of protein regions to exist as a (red) helix, (yellow) extended β-sheet or as a (green) coil, indicates the α + β components in OAC. (C) DISOPRED determines the ordered state of a protein; results indicate that OAC is an ordered protein, as is depicted by the appearance of dashes (for order) instead of asterisks (for disorder).

To identify suitable structural homologs from which an OAC structure model could be predicted, the OAC protein sequence was evaluated using the template identification feature provided by the Swiss Model Workspace (Altschul et al., 1997; Arnold et al., 2006) (Figure 6.4). Three protein structures were identified by Gapped Blast and ten protein structures were identified by HHSeach Template Library searches.

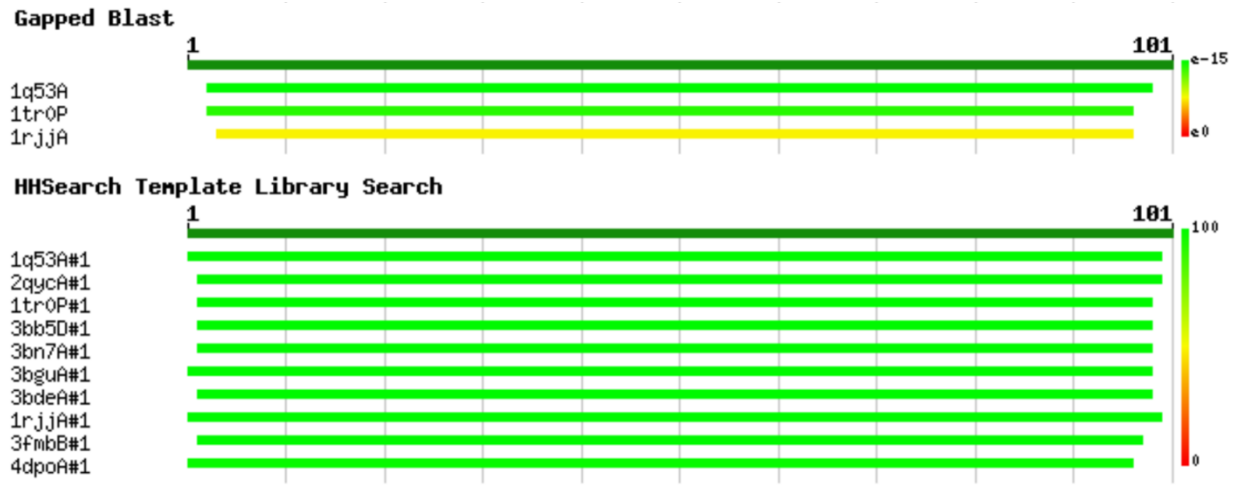


Figure 6.4 Swiss Model template retrieval of OAC.

A template search using the OAC protein sequence identified three suitable matches when a Gapped Blast was performed and ten templates when the HHSeach Template Library Search was used. Both 1Q53 and 1TR0 were each used to develop separate OAC models due to their good representation in each search.

Two homology models of OAC were created using the Swiss-Model automated service that were based on either *A. thaliana* HS1 (annotated in protein databank (pdb) 1Q53) or *P. tremula* SP1 (pdb 1TR0) (Figure 6.5). The first OAC model (OAC Model 1) used 1Q53 (chains A and B) as a template, since the template identification feature identified AtHS1 as the most probable candidate in both Gapped Blast and HHSeaches. The resulting QMEAN Z-score, an estimate of the quality of the created model, has a value of -3.25 (based on a scoring range between -4 to +4). Therefore, another model (OAC Model 2) was created using 1TR0 as a template, since it was also well represented in both searches; this second model was a good prediction according to its QMEAN Z-score of -1.2. Although both models were created using separate templates, they are visually similar. Models used two chains from each respective protein to create a homodimeric model. Each monomer contains four β-strands that align with the opposing monomer's β-strands resulting in a barrel-shaped structure at the core of the protein. Two α-helices project away from the β-sheet core in each monomer. Viewing the space-filling representation of the 1Q53 based OAC model reveals a crevice between the α-helices and the β-sheets (Figure 6.6). A similar crevice is also detected in the 1TR0 based OAC model.

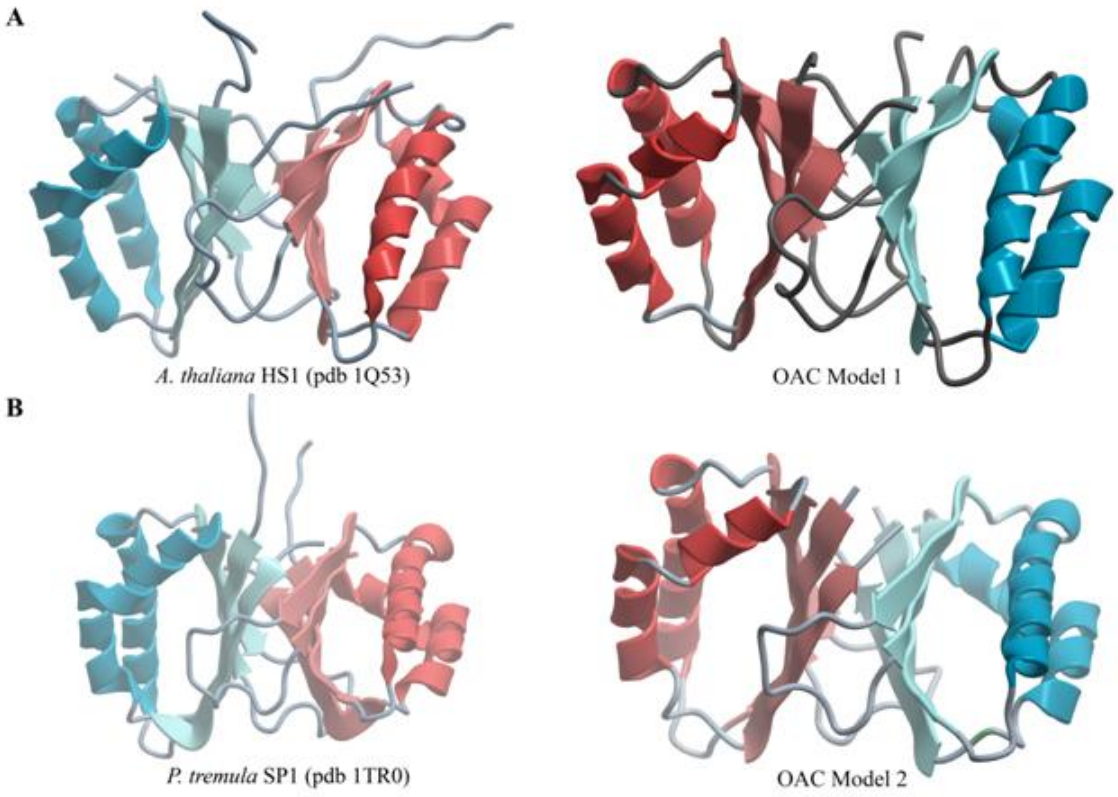


Figure 6.5 Templates and models for OAC.

(A) AtHS1 and (B) PtSP1 were each used to predict OAC models shown to the right of each template. All protein structures are shown as homodimers with separate monomers displayed in blue and red. Structures were positioned as AtHS1.

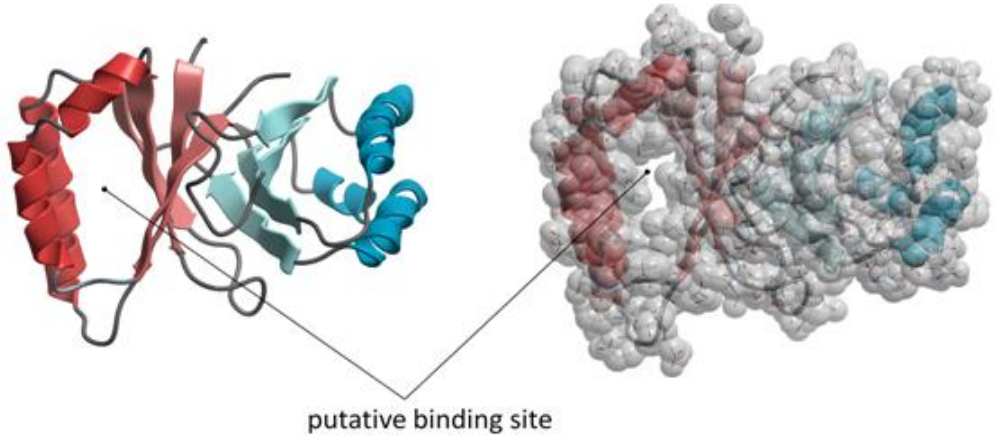


Figure 6.6 Space filled model of OAC.

OAC Model 1 from Figure 6.5A is shown tilted to highlight a crevice found between the α-helices and β-sheet, and proposed to act as the substrate-binding pocket. The model on the left is the same as the one on the right save for the addition of the space-filling features. Although not seen, a similar pocket also exists in the other monomer.

6.3.2 The Expression and Activity of OAC Mutants

The structure/function relationship of various AAs in OAC was explored by substituting alanine at AA sites conserved between SRABB homologs, in addition to His75 and His78, which are not conserved in the *P. tremula* homolog (Figure 6.2). Experiments were conducted on two occasions (experiments A and B) and were initially designed to compare the activity of 12 OAC mutants with the wild-type, but later included another three mutants (*i.e.* K4A, K38A, and D45A). The expression level of mutant proteins was determined by comparing their yields following IMAC purification to that of wtOAC (Table 6.2). Cultures expressing the Y72A, H78A, and D96A mutants in experiment A, and the D45A, H78A, and D96A mutants in experiment B produced the greatest amounts of total protein. The smallest amounts of protein were produced in cultures expressing the F11A, L30A, and G43A mutants during experiment A, and the F11A, W42A, and T56A mutants during experiment B. Cultures expressing the K12A and Y72A mutants showed the largest fluctuation in protein yields between experiments A and B; cultures expressing the K12A mutant yielded 75% less protein in those expressing wtOAC in experiment A, but no such loss was seen in experiment B; cultures expressing the Y72A mutant yielded 70% more protein in cultures expressing wtOAC during experiment A, yet yielded 7% less protein in its wtOAC counterpart during experiment B. Other than these two mutants, the amounts of protein produced by cultures showed the same trend between experiments A and B. The SDS-PAGE of wtOAC and mutants was conducted prior to experiment A (Figure 6.7). This gel does not include the K4A, K38A, and D45A mutants.

Table 6.2 OAC constructs and their yields

		Experiment A				Experiment B			
OAC version	plasmid	volume (µL)	concentration (mg/mL)	yield (µg)	expression	volume (µL)	concentration (mg/mL)	yield (µg)	expression
wt*	pJE60174	200	1.01	202	100%	100	2.50	250	100%
K4A	pJE60187	n/a	n/a	n/a	n/a	100	0.600	60	24%
H5A	pJE60175	100	0.780	78	39%	100	1.94	194	78%
F11A	pJE60176	60.0	0.430	25.8	13%	100	0.760	76	30%
K12A	pJE60177	60.0	0.850	51.0	25%	100	2.49	249	100%
L30A	pJE60178	60.0	0.400	24	12%	100	1.41	141	56%
K38A	pJE60188	n/a	n/a	n/a	n/a	100	2.22	222	89%
W42A	pJE60179	60.0	0.620	37.2	18%	100	0.850	85	34%
G43A	pJE60180	60.0	0.450	27	13%	100	1.39	139	56%
D45A	pJE60189	n/a	n/a	n/a	n/a	100	3.93	393	160%
T56A	pJE60181	100	0.330	33.0	16%	100	1.16	116	46%
H57A	pJE60182	100	0.580	58	29%	100	1.46	146	58%
Y72A	pJE60183	100	3.33	333	170%	100	2.32	232	93%
H75A	pJE60184	60.0	0.850	51	25%	100	1.47	147	59%
H78A	pJE60185	100	2.18	218	110%	100	3.42	342	140%
D96A	pJE60186	500	0.610	305	150%	100	6.39	639	260%

*wt is the wild-type OAC which did not contain any amino acid substitution. Constructs which were not included during the experiment are labelled as not applicable (n/a).

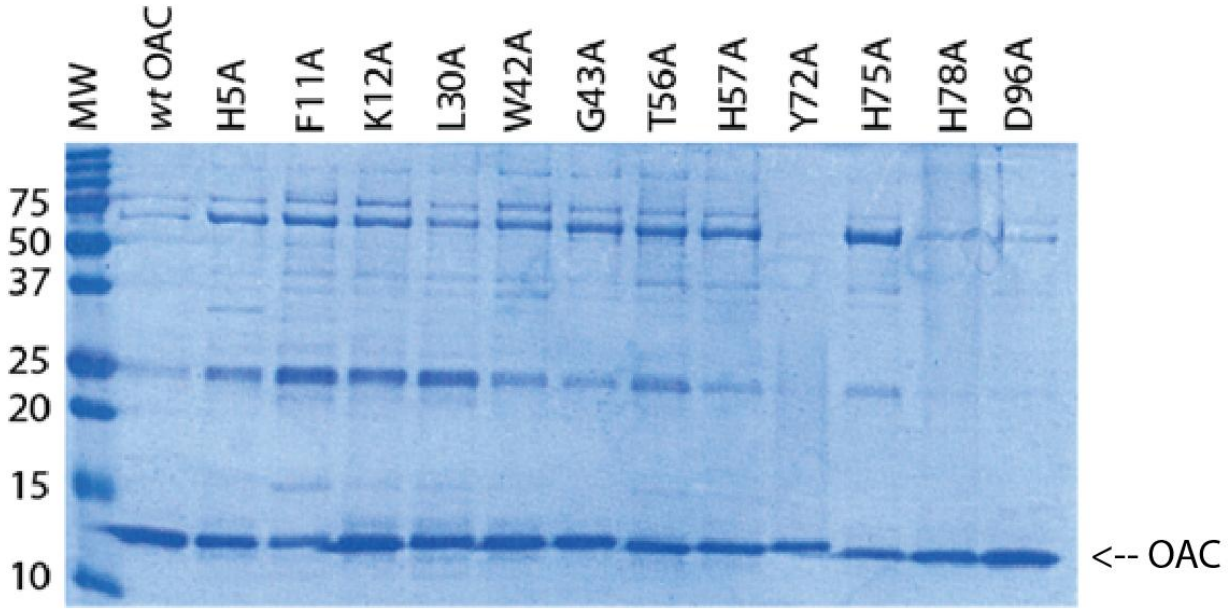


Figure 6.7 SDS-PAGE of wtOAC and mutants.

Proteins purified by IMAC were loaded onto a 15% acrylamide gel to confirm their presence and abundance. A 5 µg aliquot of wild-type OAC (wtOAC) and mutant OAC loaded on equal protein basis (names above the lanes) were run at 200 V for 45 min beside molecular weight standards (MW). Numbers on the left of the gel correspond to the kDa weight of MW. Proteins of interest are seen at the 13 kDa mark.

Although assays performed in experiments A and B are similar, they were conducted under different conditions (where the production of OA was further optimized during experiment B), and absolute amounts are therefore not averaged. Instead, results are presented separately, and the observable trends are compared. All reaction products were analyzed by LCMS, and results are reported as the total pmol yields determined by LC at 270 nm (For Experiment A see Figure 6.8; For Experiment B see Figure 6.9).

6.3.2.1 Experiment A

The activities of OAC mutants were compared to the wild-type enzyme by assaying each protein with TKS and MCS. The amount of HTAL collected during assays was between 36 ± 6 pmol and 54 ± 24 pmol (Figure 6.8A). A larger amount of variation (± 20 pmol) was detected in assays that accumulated greater than 45 pmol of HTAL, which included the L30A, W42A and D96A constructs. The largest amount of HTAL was detected from D96A assays, whereas the lowest amount was detected from G43A assays. The amount of PDAL synthesized during each assay appears static, the lowest amount was seen from G43A assays (376 ± 90 pmol); the largest amount was detected from H78A assays (484 ± 84 pmol), with the exception of the no-OAC control (580 ± 252 pmol) (Figure 6.8B). Other than wtOAC assays, which produced 167 ± 34 pmol, OA was only detected in assays performed using K12A, G43A, and D96A constructs, where they accumulated 42 ± 13 pmol, 14 ± 6 pmol, and 32 ± 14 pmol, respectively (Figure 6.8C). The standard deviation of olivetol produced during assays appear the most dynamic, where the least amount of variation was detected from assays that used wtOAC (45 ± 21 pmol), and the largest amount from L30A assays (327 ± 169 pmol) (Figure 6.8D).

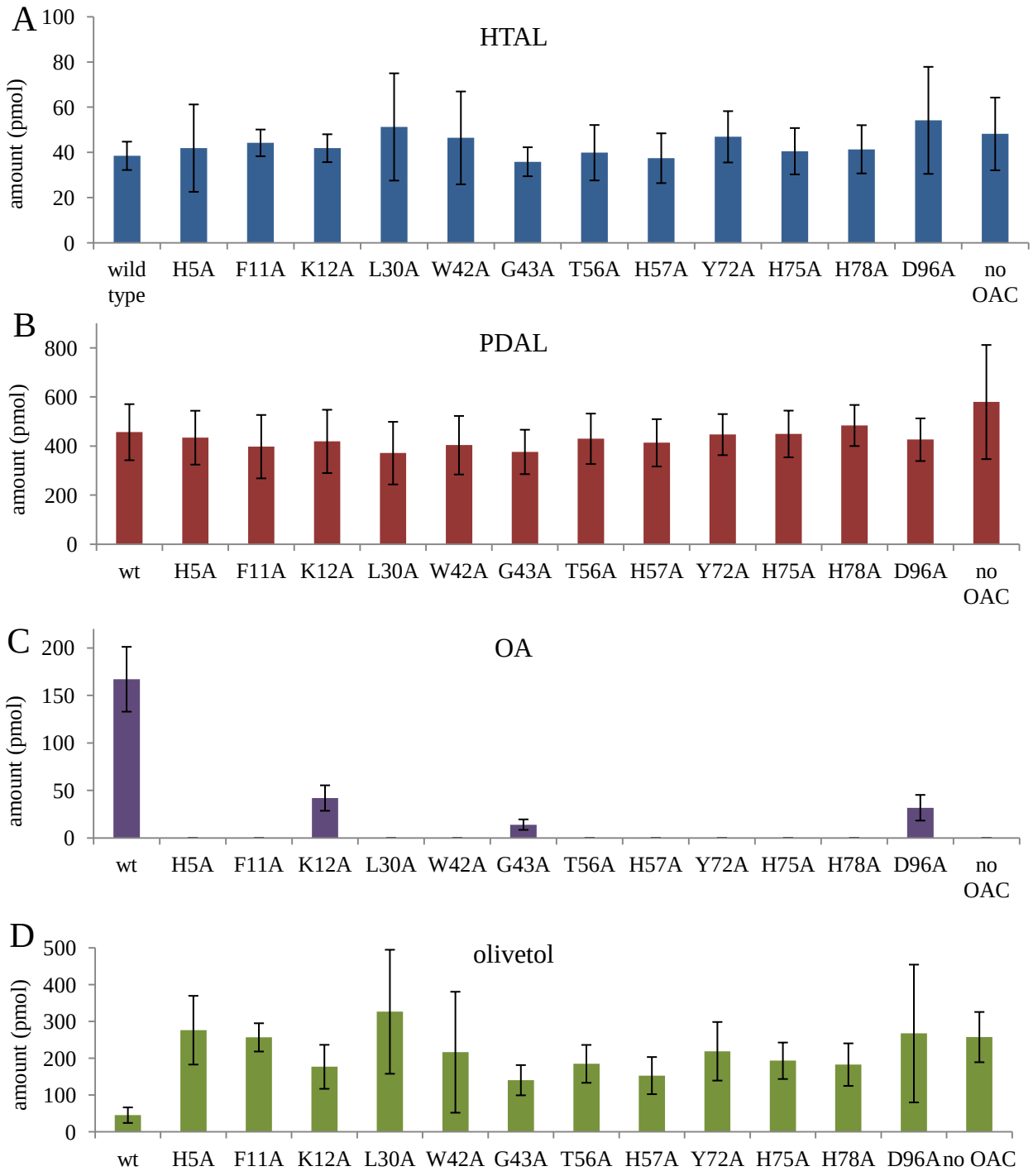


Figure 6.8 LC results for OAC mutant assays - Experiment A
Product yields for (A) HTAL, (B) PDAL, (C) OA, and (D) olivetol biosynthesized during TKS/OAC assays conducted in the presence of wild-type (wt) or mutant OAC was determined by LCMS analysis. Error bars are the SD where n=3.

6.3.2.2 Experiment B

Experimental conditions for this experiment differed from those in experiment A; conditions were optimized for greater production of OA (see Section 6.2.3.2.). Due to the number of samples tested and the necessity of processing the assays rapidly, a no-OAC control assay was not performed. All enzymes were independently expressed and purified for this experiment. The amount of HTAL collected from the assays ranged between 32 ± 7 pmol and 48 ± 16 pmol, and appeared static throughout the assays (Figure 6.9A). The largest amount of HTAL was detected from wtOAC assays, whereas the least amount was detected from both H57A and Y72A assays (Figure 6.9A). The amount of PDAL produced during each assay also appear static, where the least amount detected was 474 ± 39 pmol from D96A assays, and the largest amount was 560 ± 70 pmol during H5A assays (Figure 6.9B). Similar to experiment A, OA was detected in assays using wtOAC, K12A, G43A, and D96A constructs, where they accumulated 360 ± 134 pmol, 179 ± 16 pmol, 35 ± 13 pmol, and 23 ± 2 pmol, respectively. It was additionally detected in assays using the K4A, K38A, and D45A mutantsm, which produced 22 ± 8 pmol, 13 ± 5 pmol, 258 ± 25 pmol, respectively (Figure 6.9C). A trace amount of OA was also detected from H75A, which accumulated 4 ± 2 pmol. The dynamic behavior of olivetol production was also evident during this experiment, where the smallest amount detected was obtained from D45A, which accumulated 61 ± 18 pmol, and the largest amount was detected from H5A, which accumulated 402 ± 49 pmol (Figure 6.9D).

The amounts of OA detected during experiments A and B, as well as the relative activity compared to wtOAC are tabulated in Table 6.3. The location of amino acids targeted for SDM are shown on the OAC model designed from the 1Q53 template (Figures 6.2 and 6.10).

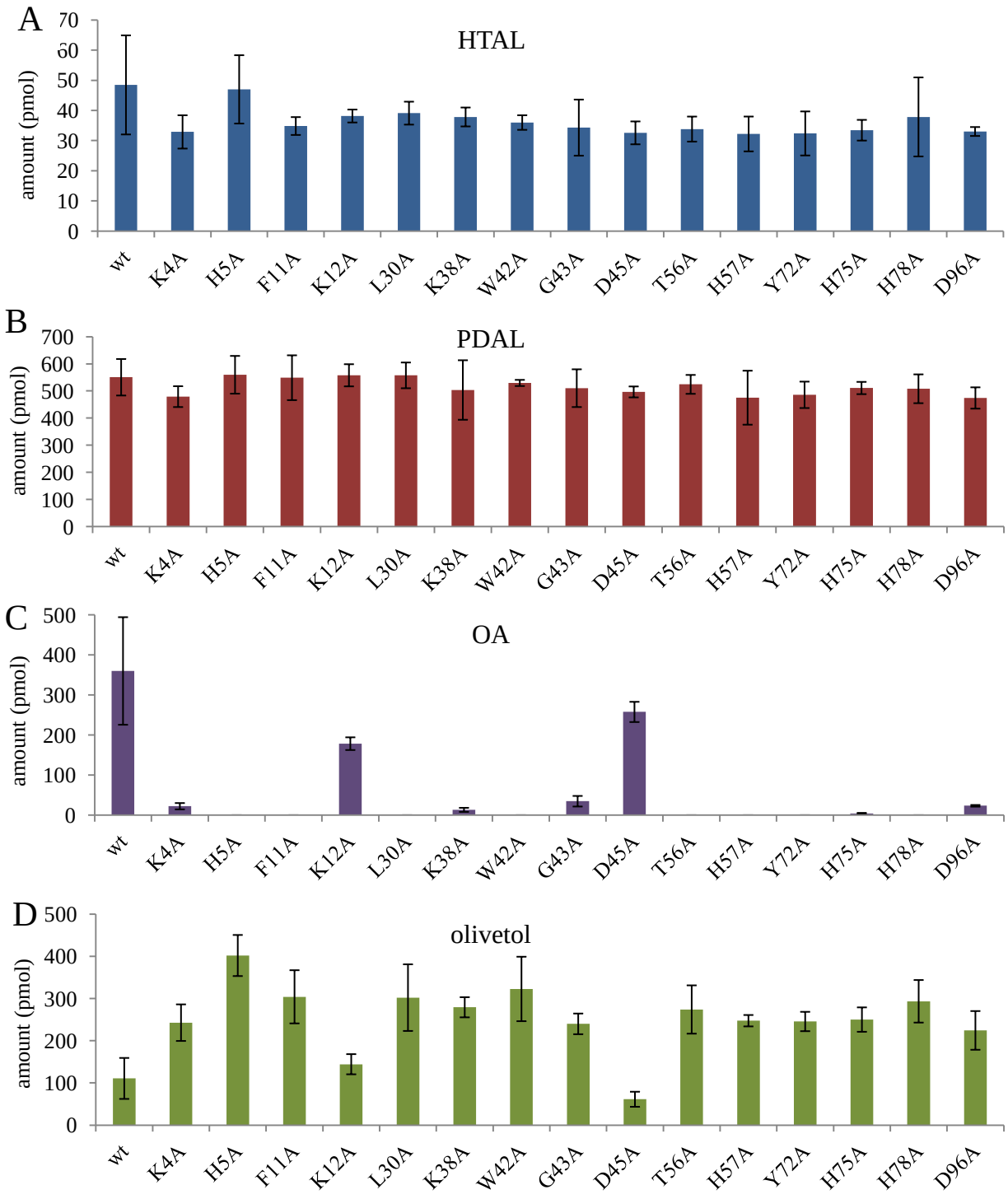


Figure 6.9 LC results for OAC mutant assays - Experiment B.

Product yields for (A) HTAL, (B) PDAL, (C) OA, and (D) olivetol biosynthesized during TKS/OAC assays conducted in the presence of wild-type (wt) and mutant OAC constructs was determined by LCMS analysis. Error bars are the SD where n=3.

Table 6.3 Successful OA biosynthesis by OAC point mutants

	Experiment A		Experiment B		
OAC version	OA (pmol)	OAC activity	OA (pmol)	OAC activity	Average relative OAC activity
wt*	167	100%	360	100%	100%
K4A	n/a	n/a	22	6.1%	n/a
K12A	42	25.1%	179	49.7%	37.4%
K38A	n/a	n/a	13	3.6%	n/a
G43A	14	8.3%	35	9.7%	9.0%
D45A	n/a	n/a	258	71.7%	n/a
H75A	0	0%	4	1.1%	0.6%
D96A	32	19.2%	23	6.4%	12.8%

*wt is the unmutated OAC. n/a samples were not tested.

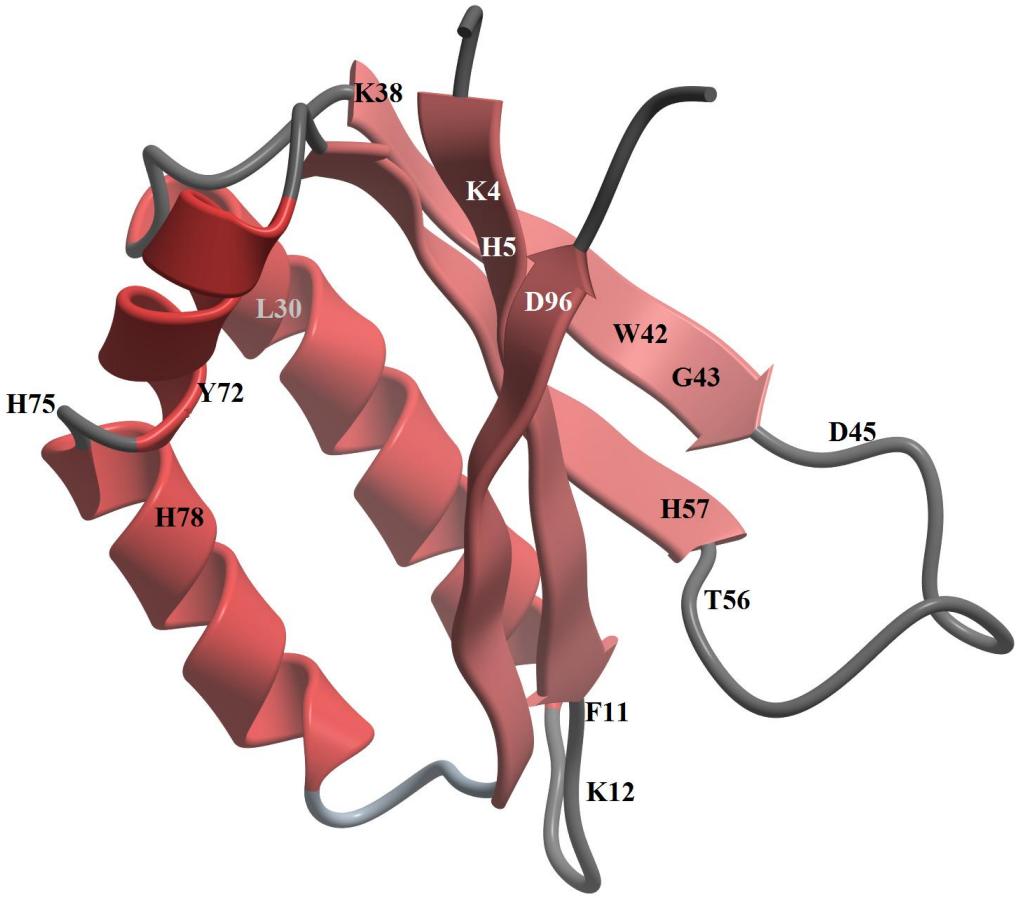


Figure 6.10 Location of SDM target sites mapped on the OAC model.

The 15 amino acids that were targeted in this study have been mapped onto the OAC model derived from the 1Q53 template. Sites are either mapped on or next to the actual location. The L30 site shown in grey exists in the α-helix drawn near the back of the figure.

6.3.3 MS/MS Monitoring of the TKS and TKS/OAC Reactions

In an attempt to detect the presence of a CoA-linked polyketide intermediate (Figure 6.1), the TKS and TKS/OAC reactions were monitored using tandem mass spectrometry. Specifically, both precursor ion scanning and selected reaction monitoring modes were used to screen the reaction mixture for CoA-linked intermediates (de Hoffmann, 1996; Glish and Vachet, 2003). Enzyme assays were analyzed using tandem mass spectrometry during multiple reaction monitoring (MRM) for parent ion and daughter ion pairs resulting from neutral loss of 507 amu, as well as 428 amu which are attributed to CoA tethered compounds of interest. Using these methods, the dissipation and generation of reaction substrates and CoA compounds was observed.

A parent ion scan was used to analyze: 1) a blank sample containing the reaction mixture but no enzymes, 2) a TKS reaction, and 3) a TKS/OAC reaction (Figure 6.11). All ions of interest included free CoA (768 m/z), acetyl-CoA (810 m/z), malonyl-CoA (854 m/z), hexanoyl-CoA (866 m/z), proposed intermediate (992 m/z), and formyl-CoA (796 m/z). Reactions were also analyzed using spectrometry in MRM mode and resulted in the same trends as those appearing during parent ion scans (Figure 6.12). Although reaction blanks did not noticeably change over time, TKS and TKS/OAC reactions accumulated large amounts of free CoA and acetyl-CoA, associated with a loss of malonyl-CoA and hexanoyl-CoA. Although not visible in Figure 6.12, trace amounts of the proposed tetraketide intermediate were detected (Table 6.4).

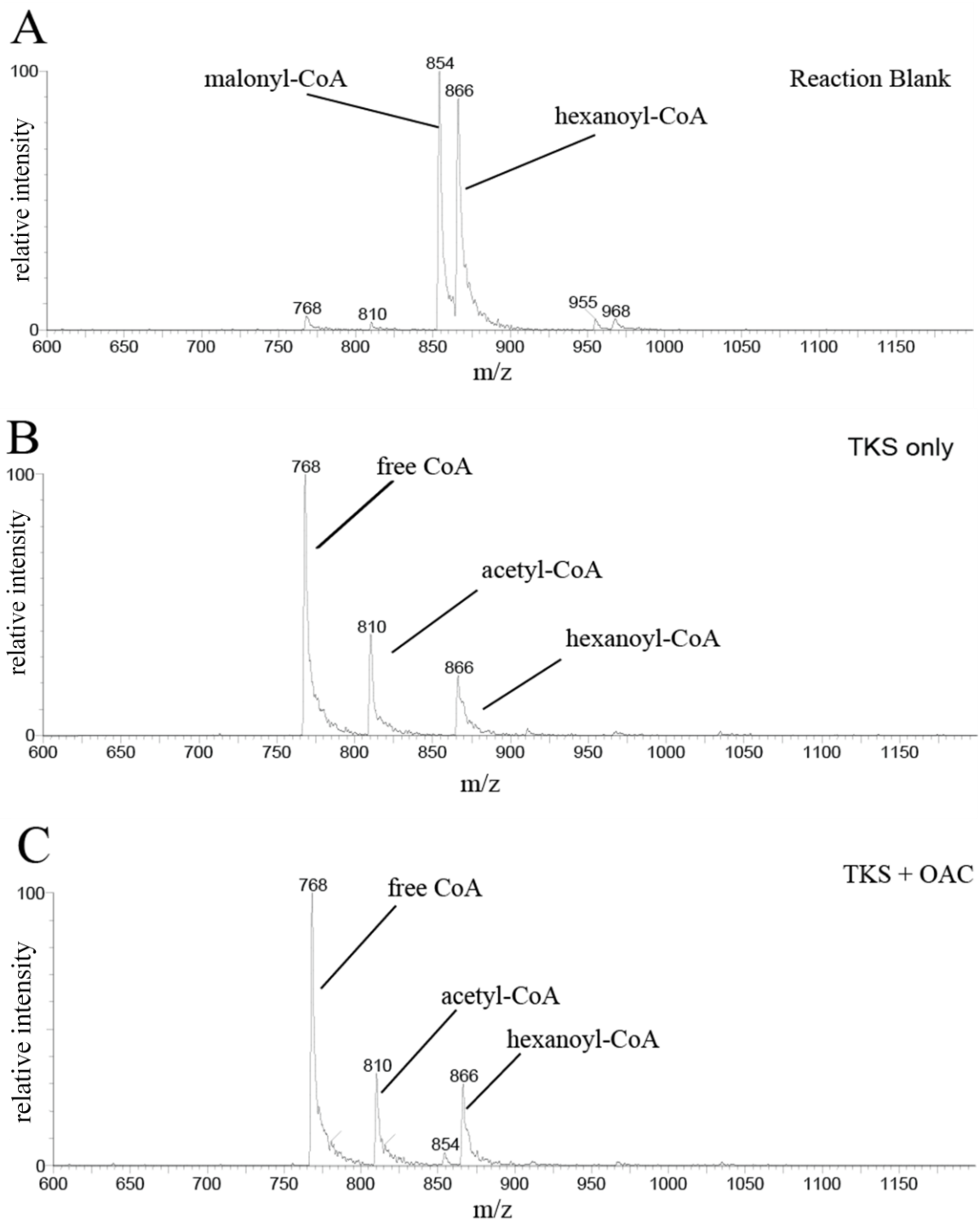


Figure 6.11 Parent ion scan of TKS and TKS/OAC reactions.

MS analysis of CoA-linked compounds was conducted on (A) a reaction blank, (B) a 60 min TKS only reaction, and (C) a 60 min TKS/OAC reaction. Parent ion scans detected fragments of 428 m/z.

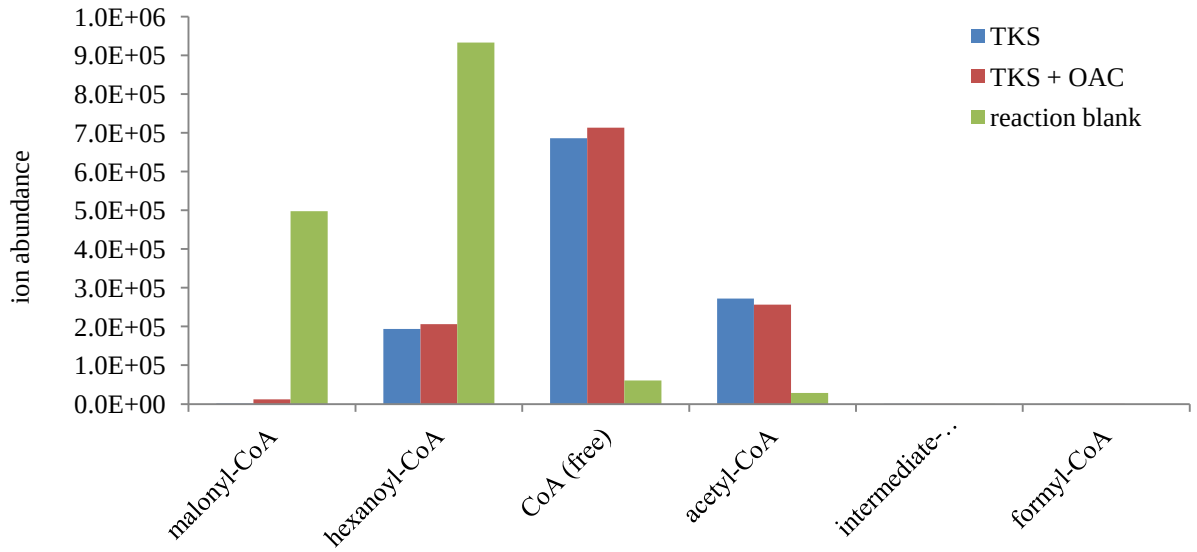


Figure 6.12 Multiple Reaction Monitoring of TKS and TKS/OAC reactions.

TKS and TKS/OAC reactions were analyzed by MS for CoA-linked compounds. Results are plotted as the ion abundance following 1 h incubation at 37°C.

Table 6.4 Area response obtained during the MRM analysis of TKS assays

	AREA					
sample	mal-CoA	hex-CoA	CoA (free)	acetyl-CoA	intermediate-CoA	formyl-CoA
TKS	1435	193633	686457	272484	449	9
TKS + OAC	12047	206387	713452	256350	471	11
reaction blank	497697	933147	60587	28518	4	0

A TKS reaction and sample blank were also monitored for the neutral loss of CoA (507 m/z), resulting in the same trends detected previously during MS analyses (Figure 6.13). Compared to sample blanks, TKS assays showed a 60% reduction in hexanoyl-CoA, and a 99% reduction in malonyl-CoA associated with an increase of free CoA and acetyl-CoA. Neither a tetraketide intermediate nor a hypothetical formyl-CoA byproduct were detected during MRM screens.

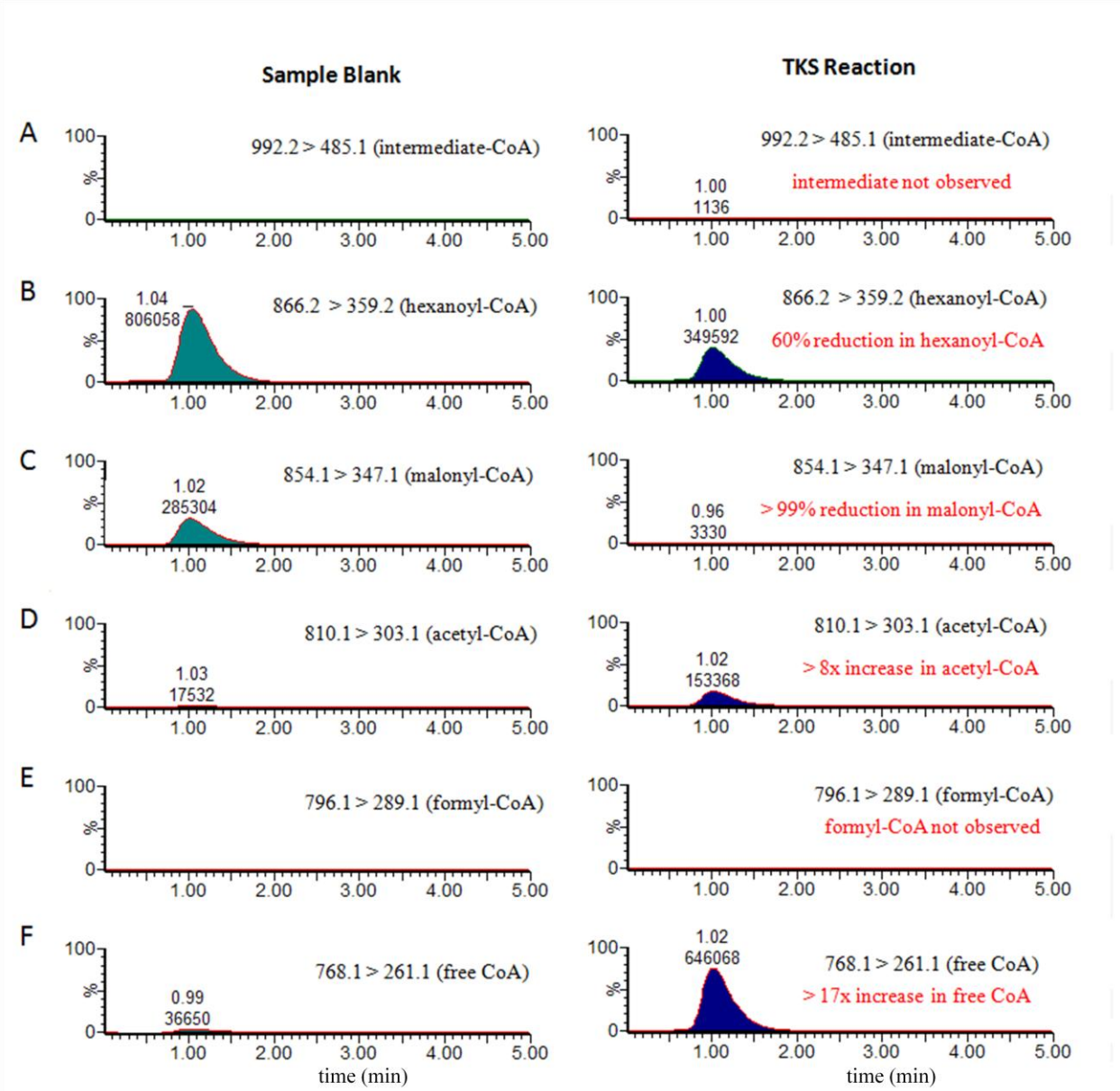


Figure 6.13 MRM for the neutral loss of CoA during a TKS reaction.

Tandem mass spectrometry was used to analyze (left side) a sample blank containing no enzyme, and (right side) a TKS reaction by screening for daughter ion pairs resulting from the neutral loss of CoA fragments (507 m/z). Samples were screened for (A) a tetraketide CoA intermediate, (B) hexanoyl-CoA, (C) malonyl-CoA, (D) acetyl-CoA, (E) formyl-CoA, and (F) free CoA. Results are plotted as the % relative intensity relative to time (min).

6.4 Discussion

6.4.1 Primary and Secondary Structure

Results indicate that OAC is a DABB protein belonging to the SRABB family. Repeated BLAST searches using the OAC primary sequence suggest a close relationship to several plant homologs. These homologs were used to create an alignment that led to the identification of conserved amino acids constituting an overall 18% coverage of the SRABB homologs. The high degree of sequence similarity suggests a strong relationship exists between structure and function, likely due to the smaller size of the proteins, resulting in the retention of their specific shape, which is required for activity. OAC models predicted from AtHS1 and PtSP1 templates confirm that OAC is a DABB protein made up of two alternating βαβ motifs. Specifically, the βαβ motifs appear more like a βαββααβ, but note that the repeated αα helices are really one α helix that have adopted a kink halfway through its span. This kink provides a necessary space between the α-helices and β-sheet previously proposed and shown to bind substrates in a number of other DABB proteins including *S. coelicolor* ActVa-Orf6, *S. galilaeus* AknX, and *S. aureus* IsdG (Chung et al., 2002; Lee et al., 2008; Sciara et al., 2003).

The 1Q53 and 1TR0 structures served as appropriate templates to model an OAC structure. Lower-quality modeling using the 1Q53 template may have been due to the NMR method of deduction used to solve its structure, which may have resulted in the exclusion of some functions that the modelling software requires, since the software was primarily designed for structures obtained by crystallography. The OAC model developed from the 1TR0 template resulted in a similar OAC structure obtained from 1Q53, although less error was produced possibly due to the methods of crystal elucidation used for 1TR0.

6.4.2 Site-Directed Mutagenesis

SDM-targeted amino acids, in most cases, resulted in the abolishment of OAC activity. Amino acid targets themselves were based on their conservation within OAC homologs, but also placed some priority on establishing whether or not they played a critical role in OAC activity. Some amino acids were not suspected to play any role in activity, such as G43, yet their substitution resulted in a dramatic loss of activity. Again, this suggests that a strong relationship exists between structure and function in these proteins, probably due to its smaller size, resulting in misfolding of mutant proteins upon their expression. The inclusion of K4, K38, and D45 during the second round of experiments was an attempt to delineate their potential roles in either Schiff base formation, which would implicate a class I aldolase mechanism, or through an acid/base-catalyzed reaction, which could involve an aspartate residue. The substitution of these residues, as well as with K12 and D96, did not result in the total loss of function, which would have otherwise suggested their direct involvement in the OAC mechanism. Instead, the K4A, K38A, and D45A mutants, apart from G43A and H75A, were the only mutants that retained any OAC activity, whereas all other mutants failed to produce OA when assayed with TKS.

The identification of OAC as a putative cyclase was originally based on its structural similarity with another class of DABB proteins, the PKCs. The PKC family is exemplified by TcmI, the only characterized PKC to date. TcmI is part of the tetracenomycin biosynthesis gene cluster found in *Streptomyces glaucescens* and was shown to cyclize Tcm F2 to Tcm F1, both intermediates involved in the biosynthesis of the antitumor antibiotic Tcm C, without requiring a cofactor or metal ion (Shen and Hutchinson, 1993b). Since the structural elucidation of TcmI lacks an enzyme-substrate complex, the active site was inferred to exist at an analogous α/β interface as proposed in ActVa-Orf6 (Sciara et al., 2003). In TcmI, the proposed active site is

lined with many hydrophobic residues conducive to binding polyaromatic compounds, where a single sulfate ion was shown bound to R40 and H26 at the back of the pocket. TcmI residues (H26, D27, R40, and H51) that were putatively involved in catalysis were mutagenized, but results indicate the residues are most likely involved in hydrogen bonding or polar interactions and are not critical for enzyme activity. Although no specific mechanism was offered by Thompson et al. (2004), they suggested that the role of TcmI was to position the substrate in a way which accommodates for a water-facilitated reaction. I also postulate that the active site for OAC exists in the analogous α/β interface proposed in other DABB proteins; similar to TcmI, this interface also contains many hydrophobic residues that could potentially bind and aid in the cyclization of a tetraketide intermediate (Figure 6.14).

The intramolecular aldol condensation performed by TcmI is also reminiscent of similar cyclization reactions mediated by SnoaL-like polyketide cyclases (interpro domain IPR009959), which also require neither Schiff-base formation nor a metal ion (Sultana et al., 2004). Although not structurally related, *S. nogalater* SnoaL is a polyketide cyclase involved in the biosynthesis of nogalaviketone, where a conserved D121 is thought to act as an acid/base catalyst. Notably, this distinct SnoaL motif is sometimes found fused to ABM members and is also found upstream from TcmI, where it is involved in an earlier polyketide cyclization. SRABB homologs, including OAC, also contain a conserved aspartate residue (D96) analogous to the D121 residue found in SnoaL. The D96A mutant was created to determine whether or not this residue was involved in an acid/base mechanism critical for OAC activity, however assay results disqualify this scenario. Based on its location and assay results, this residue may instead be involved in either attracting the substrate to the active site cavity, or the stabilization of the dimer interface.

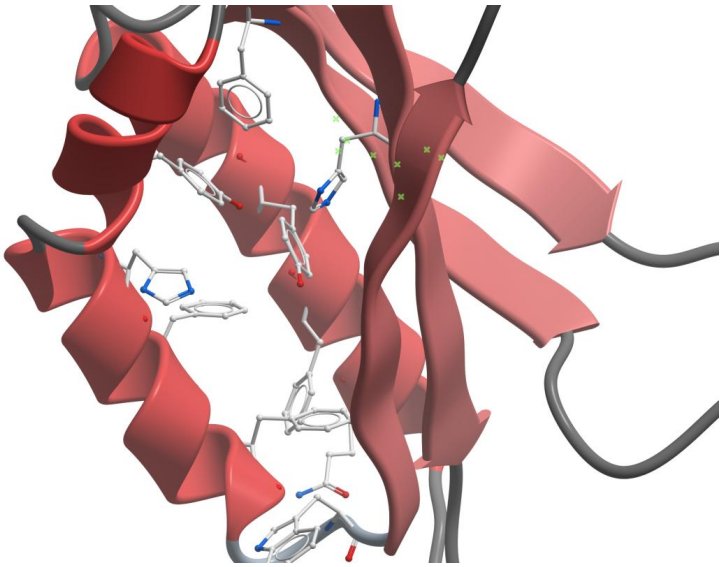


Figure 6.14 The proposed active site for OAC.

The proposed active site found between the α/β interface is lined with many hydrophobic residues conducive to the potential binding of tetraketide intermediates. These amino acids also include two tyrosine residues (oxygen marked in red) and two histidine residues (nitrogen marked in blue) that could assist in cyclization.

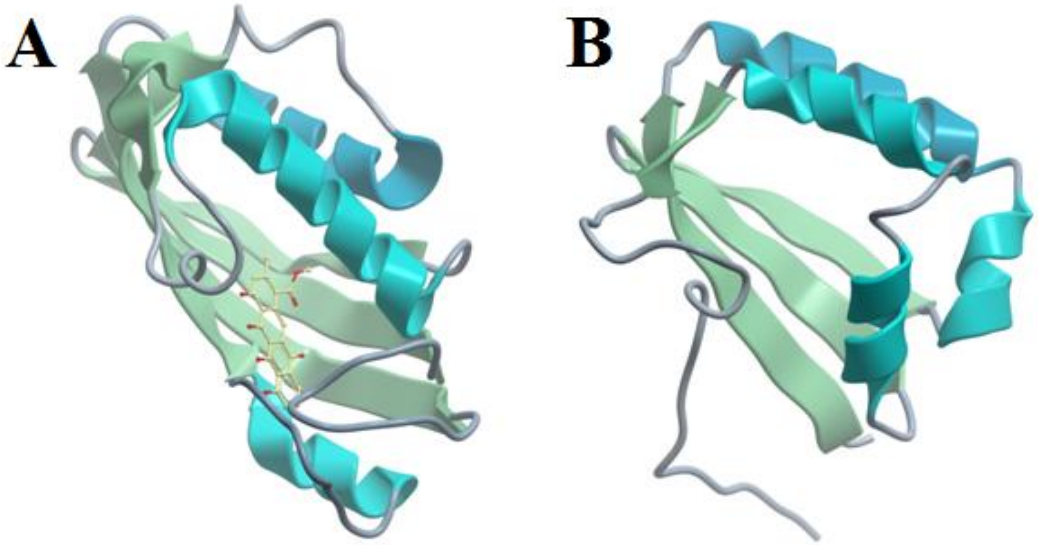


Figure 6.15 The structure of SnoaL and TcmI.

(A) The structure for SnoaL (pdb 1SJW) has been shown to bind substrates between a β-sheet and its α-helices, which has also been proposed to occur in (B) TcmI (pdb 1TUW), a PKC belonging to a different family of proteins than SnoaL. Both enzymes perform aldol-like reactions on polyketide-derived intermediates.

6.4.3 CoA analysis of the TKS reaction

The MS/MS MRM analyses were conducted to detect CoA-linked intermediates which could help resolve TKS and OAC reaction mechanisms. Reactions were specifically monitored for the presence of TKS substrates, as well as for putative reaction products and intermediates including free CoA, acetyl-CoA, and a hypothetical tetraketo-CoA ester. Experiments also monitored formyl-CoA, which could have indicated an exit strategy for CO_2 during the formation of olivetol, but this hypothetical model was abandoned after no formyl-CoA was detected. All MS/MS analyses indicate that the TKS reactions proceeded until malonyl-CoA, but not hexanoyl-CoA, was depleted. The analyses also show that acetyl-CoA is produced during TKS reactions. This is likely due to the escape of enol intermediates from the TKS active site following the decarboxylation of malonyl-CoA. Normally, this decarboxylation drives the addition of acetyl groups onto a nascent polyketide (Austin and Noel, 2003), but suboptimal *in vitro* conditions could result in the diffusion of labile ethenol-CoA esters into the surrounding medium. The detection of acetyl-CoA may also raise questions as to why it is not being used to produce truncated methyl resorcinol/ate analogs such as those observed during TKS assays that used acetyl-CoA as a starter compound (Section 4.2.10.5), but this growing acetyl-CoA pool would be in direct competition with hexanoyl-CoA in an ever dwindling pool of malonyl-CoA.

Although MRM for neutral loss of CoA was incapable of detecting any polyketide intermediate (992 m/z), MS analyses using a SIR of 428 m/z successfully detected trace amounts of a putative tetraketo-CoA ester during TKS and TKS/OAC reaction monitoring, which notably was not detected in sample blanks. Although the trace detection of this intermediate using these methods could be credited as a success, results are inconclusive. The detection of a tetraketo CoA-ester may be limited due to its lability under harsh conditions created during MS ionization.

7. GENERAL DISCUSSION

7.1 Introduction

Throughout this study TKS assays conducted in the presence of malonyl-CoA and hexanoyl-CoA have repeatedly produced HTAL, PDAL and olivetol. The addition of OAC to this reaction resulted in the additional biosynthesis of OA. The results have also shown that neither TKS nor OAC are capable of producing OA without each other's activity. As well, TKS and OAC do not appear to physically interact, and OAC retains activity even when TKS was removed from solution, therefore dismissing the notion that an allosteric relationship is involved between TKS and OAC (Chapter 5). Additionally, results have indicated that TKS and OAC do not require a metal, or Schiff base formation, dismissing the traditional class I and class II aldolase mechanisms. Instead OAC activity is reminiscent of other unrelated polyketide cyclases, namely SnoaL, which also does not rely on class I or class II aldolase mechanisms to achieve cyclization. Instead, these results suggest that an intermediate is produced by TKS, which is released into the medium and then cyclized to OA by OAC. It was demonstrated that, although TKS produces HTAL, PDAL and olivetol, these compounds are unlikely the substrate for OAC (Chapter 4). Attempts to detect an intermediate were equivocal, yet trace amounts of intermediate were detectable in TKS and TKS/OAC assays when they were analyzed by tandem mass spectrometry (Table 6.4), which were not present in reaction controls (Figure 6.13).

Although my observations do not provide conclusive evidence on either the TKS or OAC mechanism, this study does provide the first evidence allowing model mechanisms to be hypothesized. The simple most important question to address is waht is the substrate for OAC; o sot a CoA-linked intermediate, or a free tetraketoacid? This question will be discussed in the

following section, first addressing the TKS reaction and its products, and then the OAC reaction and how it utilizes the TKS product.

7.2 A Mechanistic View on the Synthesis of a Linear Tetraketide

TKS is involved in the production of not only a resorcinol, but also that of two lactone products, HTAL and PDAL. The biosynthesis of lactones by type III PKS is not a new observation; many instances of lactone production have been documented, where PKSs that are putatively involved in the biosynthesis of natural products yield lactone products instead of the proposed end products (Figure 2.7). As well, the production of lactones by PKSs is not only exclusive to type III PKSs. It has been reported that the reconstruction of the modular type I PKS, DEBS, and the following repositioning of the thioesterase domain from the terminus of module 6 to a preceding module 2 resulted in the interruption of chain elongation and premature release of triketide lactones (Staunton and Weissman, 2001). In their review, Staunton and Weissman (2001) also note that many α-pyrones are produced during experiments using minimal or depleted PKSs, and remark that the formation of these α-pyrones could provide a favorable mechanism for product release. In contrast to type I PKSs, type III PKSs do not possess a transesterase domain, and the issue of product release remains somewhat obscure (Austin and Noel, 2003; Jez et al., 2000). It is my proposal that with regard to TKS, and in agreement with Staunton and Weismann, lactones are artifacts of premature off-loading, but can also result during the re-loading of the active site with a tetraketide that is incapable of further extension due to constraints in the active site. Specifically, PDAL may be due to the premature off-loading of an enzyme-bound triketide. The loss of a hydrogen at the C_4 position results in enolate formation. This leads to a $C_5O \rightarrow C_1$ attack, resulting in the formation of PDAL, as well as its release from the enzyme (Figure 7.1). On the other hand, HTAL biosynthesis results following the reattachment of a tetraketide to the

active site cysteine. This intermediate can either be released by hydrolysis as a free acid, or by cyclization to HTAL following a $C_5O \rightarrow C_1$ attack (Figure 7.1). This proposal has been written with consideration to current type III PKS models explaining polyketide biosynthetic activity, as proposed by Jez et al (2000), who successfully elucidated the active site for 2–pyrone synthase in *Gerbera hybrida*.

The proposed TKS mechanism involves the release of a CoA-tethered linear tetraketide. This is an important conclusion which is supported by several observations. Firstly, product ratio comparison between TKS assays conducted with or without OAC show that OA is produced at the cost of olivetol production, but that HTAL production remains the same whether OAC is present or not (Figure 4.7). The constant amount of HTAL synthesized indicates that, although a common intermediate exists between olivetol and OA, HTAL production appears independent. This phenomenon is explained using my proposed mechanism, which shows that HTAL results from the reattachment of a CoA-tethered tetraketide to the TKS active site instead of its diffusion into the surrounding solution (Figure 7.1). Additionally, studies by Harris and Carney (1966, 1967) that focused on the cyclization of polyketoacids (and not their CoA analogs) failed to report the detection of any lactones during their experiments, indicating that pyrones cannot be spontaneously formed from free polyketoacids and that their production results from either CoA polyketoesters or, as I propose, due to enzyme binding of CoA-esters (Harris and Carney, 1966; Harris and Carney, 1967). Factors that markedly affected lactone production included starter compound availability (Chapter 3), which would directly affect competition for TKS active site attachment and result in less HTAL production, as more hexanoyl-CoA becomes available, and the adhering of TKS to Talon, which could potentially help to stabilize the movement and architecture of the TKS active site during *in vitro* reaction.

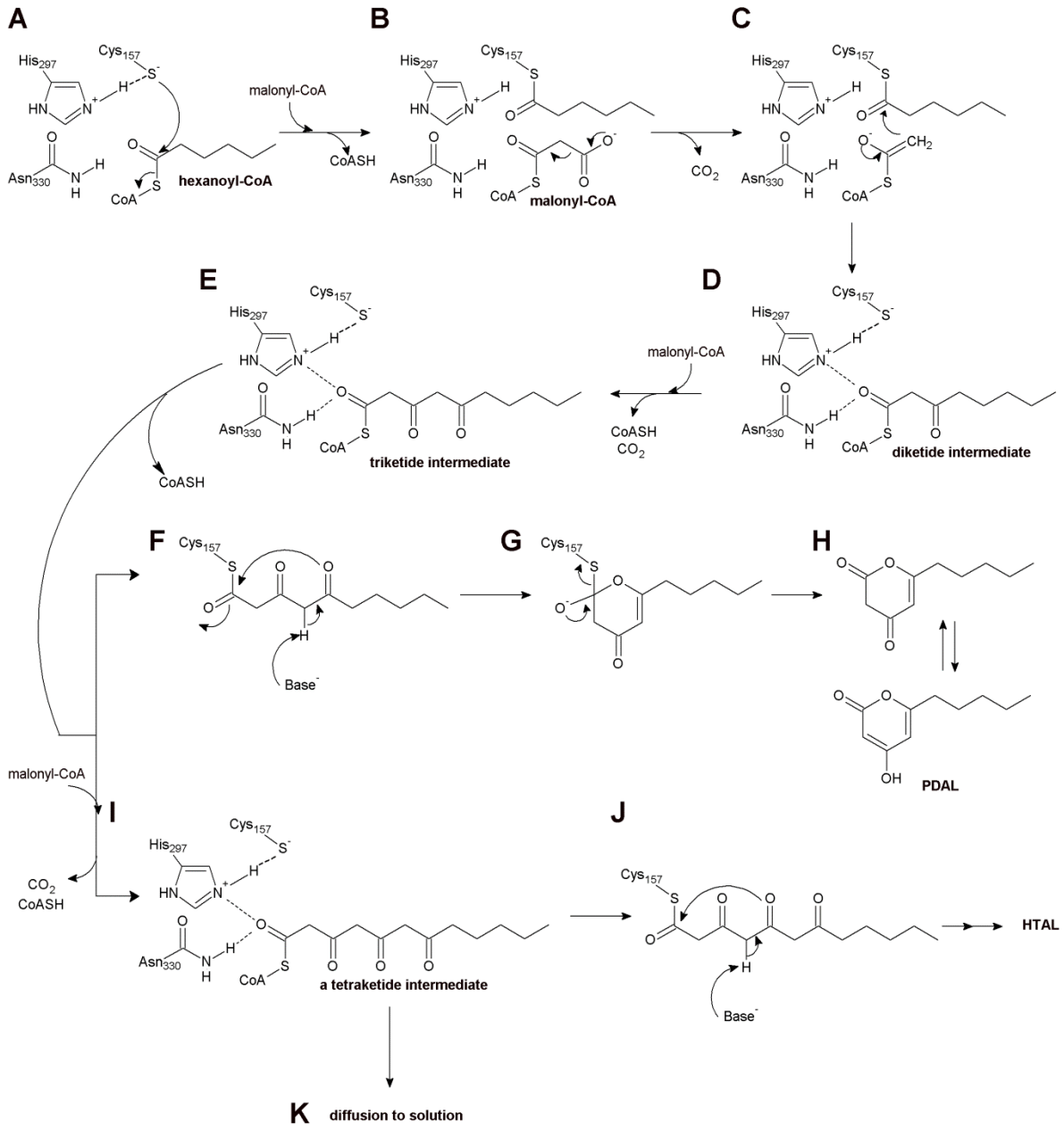


Figure 7.1 A proposed mechanism for TKS.

(A) A starter molecule is attached to the TKS active site Cys157 resulting in the loss of CoA. (B) Acetyl-CoA is activated following the decarboxylation of malonyl-CoA, (C) and attacks the starter's C1 carbonyl (D) releasing the starter from Cys157 and resulting in the formation of diketide. (E) This process is repeated by using the extended polyketide as a starter compound, resulting in the formation of a triketide. Once the triketide is reattached to the enzyme, the intermediate may either (F) spontaneously cyclize following the abstraction of hydrogen and a C5O → C1 attack, (G) releasing the lactone from Cys157, and (H) forming PDAL by tautomerization; or (I) be extended by another two carbons forming a tetraketide, that may either be: (J) reattached to the enzyme, and form a lactone (F); or (K) diffuse to the environment. A portion of this figure was adapted from Austin and Noel (2003).

7.3 The Biosynthesis of Olivetol and OA

If there does exist a diffusible, labile tetraketo-CoA ester, the mechanism for olivetol and OA biosynthesis would still require the displacement of the CoA moiety. Olivetol biosynthesis itself requires the hydrolysis of the CoA thioester, as well as a decarboxylation and, although we have seen that the decarboxylation of cannabinoid acids occurs at higher temperatures and over extended storage times, the *in vitro* decarboxylation of any polyketide in solution would drive further chemical modification, such as cyclization. The biosynthesis of OA and olivetol are substantially different in that the biosynthesis of olivetol is preceded by the hydrolysis of CoA, whereas the biosynthesis of OA requires the retention of the CoA ester, which is hydrolyzed only after the cyclization reaction is complete.

7.3.1 The Biosynthesis of Olivetol

The biosynthesis of olivetol requires a free tetraketoacid in solution, and is preceded by the hydrolysis of the CoA ester prior to decarboxylation and cyclization of the tetraketoacid (Figure 7.2). Following hydrolysis, the loss of the acidic hydrogen from the carboxyl group subsequently results in the decarboxylation of the tetraketoacid resulting in the loss of C1. The decarboxylation results in the production of a short-lived enolate. The enolate is acts as a strong nucleophile allowing a C2 → C7 nucleophilic attack to proceed, resulting in the cyclization of the intermediate to a phenolic alcohol. The phenolic alcohol is dehydrated, which results in the biosynthesis of the tautomeric form of olivetol. Note that the hydrolysis of the CoA-tetraketoacid leads to its spontaneous decarboxylation, which ultimately drives the cyclization to the final product.

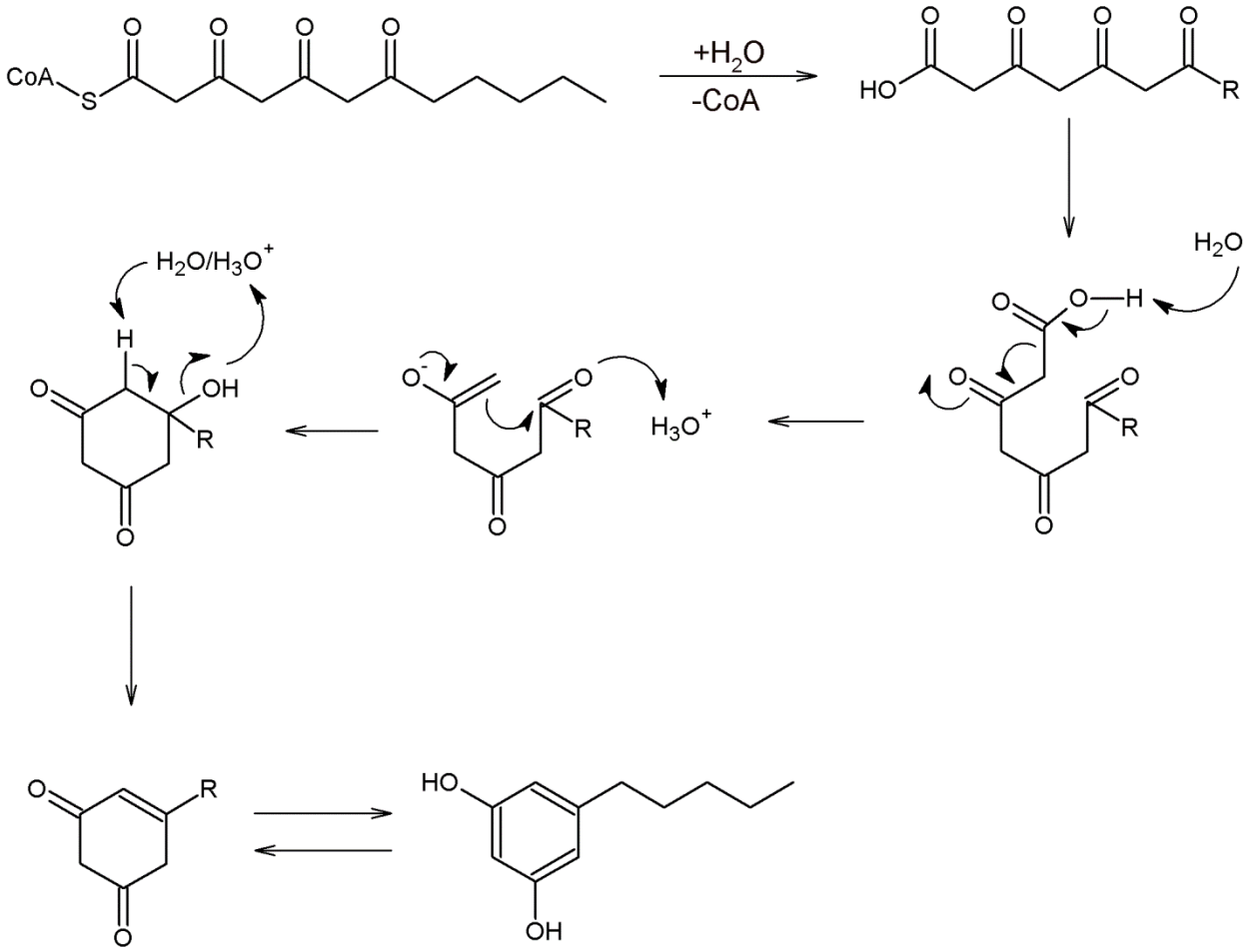


Figure 7.2 A proposed reaction mechanism for the synthesis of olivetol.

In this scheme, the R-substitutent is representing a pentyl chain, but may also occur in nature as a shorter (propyl) or longer (heptyl) alkyl chain.

7.3.2 The Biosynthesis of OA

I proposed that the biosynthesis of OA requires a tetraketo-CoA ester, which makes the tetraketide intermediate resilient to decarboxylation (Figure 7.3). Since CoA acts as a protecting group which is itself susceptible to hydrolysis in aqueous solutions, the biosynthesis of OA also requires OAC to anchor and retain the CoA ester during OA cyclization. The biosynthesis of OA proceeds once the C2 α-hydrogen is abstracted by a base, resulting in the formation of an enolate at the C1 position. This nucleophile does not hydrolyze CoA due to its position within OAC. This enolate drives the C2 → C7 cyclization, resulting in the formation of an alkoxide. The alkoxide abstracts hydrogen (presumably from the acid/base that is initially involved) to form an alcohol. Another α-hydrogen is abstracted from the C2 position, leading to the formation of the C3 enolate, which is followed by the protonation of the alcohol (again from the acid/base that is initially involved). The intermediate is then isomerized and dehydrated, resulting in the synthesis of a CoA-tethered OA tautomer. The ester is hydrolyzed, liberating free CoA and OA (Figure 7.3).

This model for OA biosynthesis shows why OAC is required for proper cyclization, since without this enzyme, tetraketo-CoA esters would simply hydrolyze and lead to the spontaneous cyclization of olivetol. Classical aldol-type mechanisms for OAC may be excluded based on the observations that alanine substitution of lysine residues only resulted in decreased OAC activity and that OAC does not require a metal. Notably, OAC mutants harboring alanine substitutions at histidine residues positioned at 5, 75, and 78, which are all located near the proposed active site, result in the abolishment of OAC activity. Given that optimum OAC activity occurs at a pH of 6.4 (Chapter 4), and that the average pKa of hisitidine is 6.5, it seems likely that these histidine residues play an important part during the acid/base reactions required to drive OA biosynthesis.

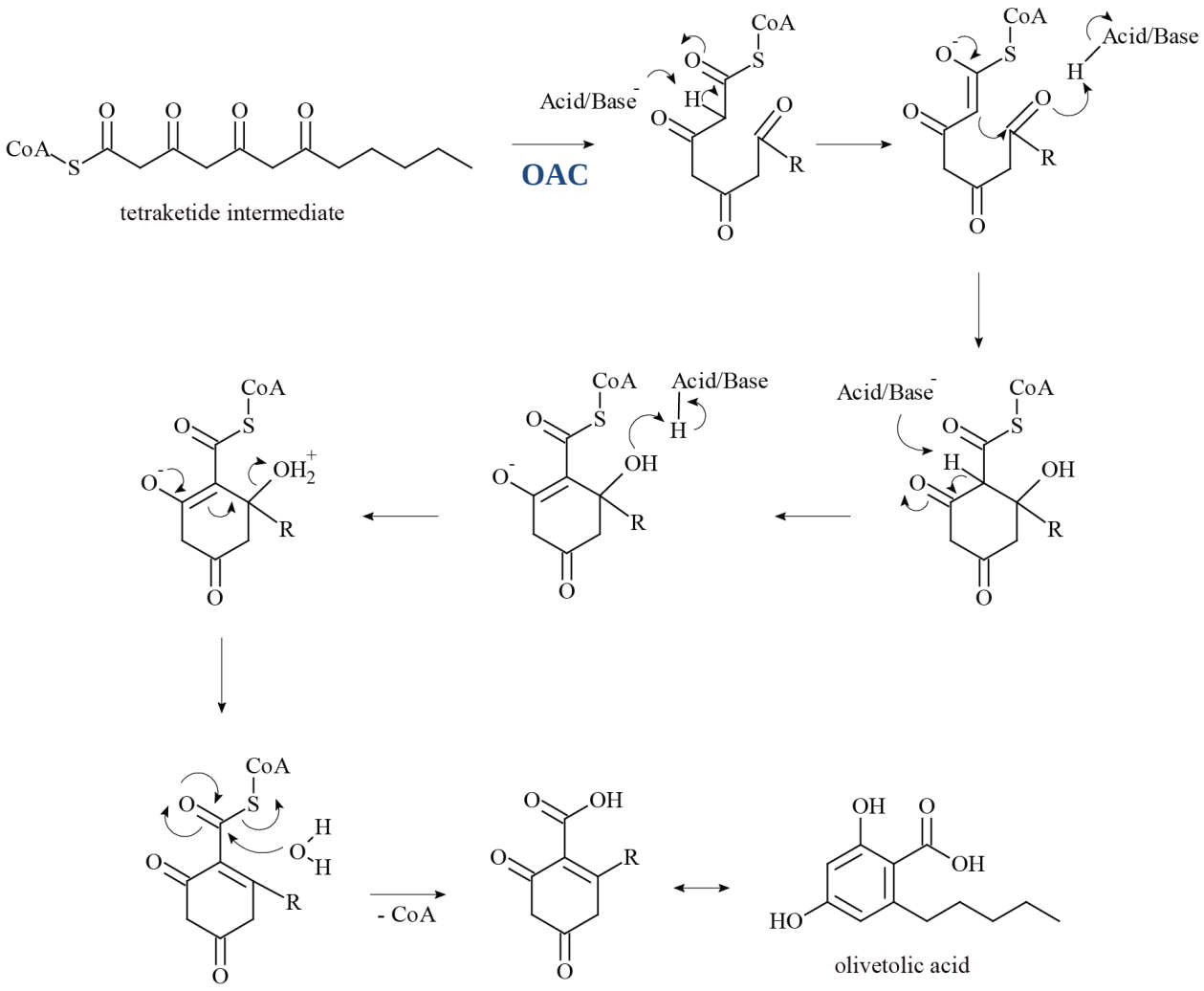


Figure 7.3 A proposed reaction mechanism for the biosynthesis of OA.

The biosynthesis of OA is preceded by the attachement of a tetraketo-CoA ester to OAC.

7.4 Cannabinoid Biosynthesis

A revised biosynthetic pathway for all cannabinoids produced in cannabis is here proposed. Although the discovery of oxidoreductases involved in the cyclization of CBGA have contributed to the delineation of later cannabinoid biosynthesis, including the biosynthesis of THCA by THCA synthase (Taura et al., 1995), and for CBDA by CBDA synthase (Taura et al., 1996), the earlier step involving OA biosynthesis has eluded description due to the lack of a PKS capable of independently producing OA. The recruitment of OAC during this preliminary step allows OA to be successfully biosynthesized, and together with TKS, the inclusion of OAC clarifies the cannabinoid biosynthetic pathway (Figure 7.4). Given its activity and abundance in trichome cDNA libraries, OAC is a key enzyme in the cannabinoid biosynthetic pathway.

7.5 Metabolic Engineering of Cannabinoid Biosynthesis

The engineering of secondary plant metabolites is an important field that has provided economically important chemicals to be produced in higher yields that are more financially worthwhile (Verpoorte et al., 2007). The discovery of OAC may play a role in the future metabolic engineering of cannabis to either augment or abolish cannabinoid content. These achievements would result in cannabis strains that have either increased medicinal value or, conversely, being devoid of cannabinoids, in an agricultural benefit due to the illegality of psychotropic cannabis. Whether it is used to increase or decrease cannabinoid content, OAC will undoubtedly play a role in the future engineering of cannabis. A practical application of OAC during *in vitro* syntheses will also benefit the future development of synthetic cannabinoid biosynthesis. The *in vitro* biosynthesis of cannabinoids has been hampered in part by the lack of OA production; the inclusion of OAC could ameliorate this problem by providing a suitable substrate for GOT to produce CBGA.

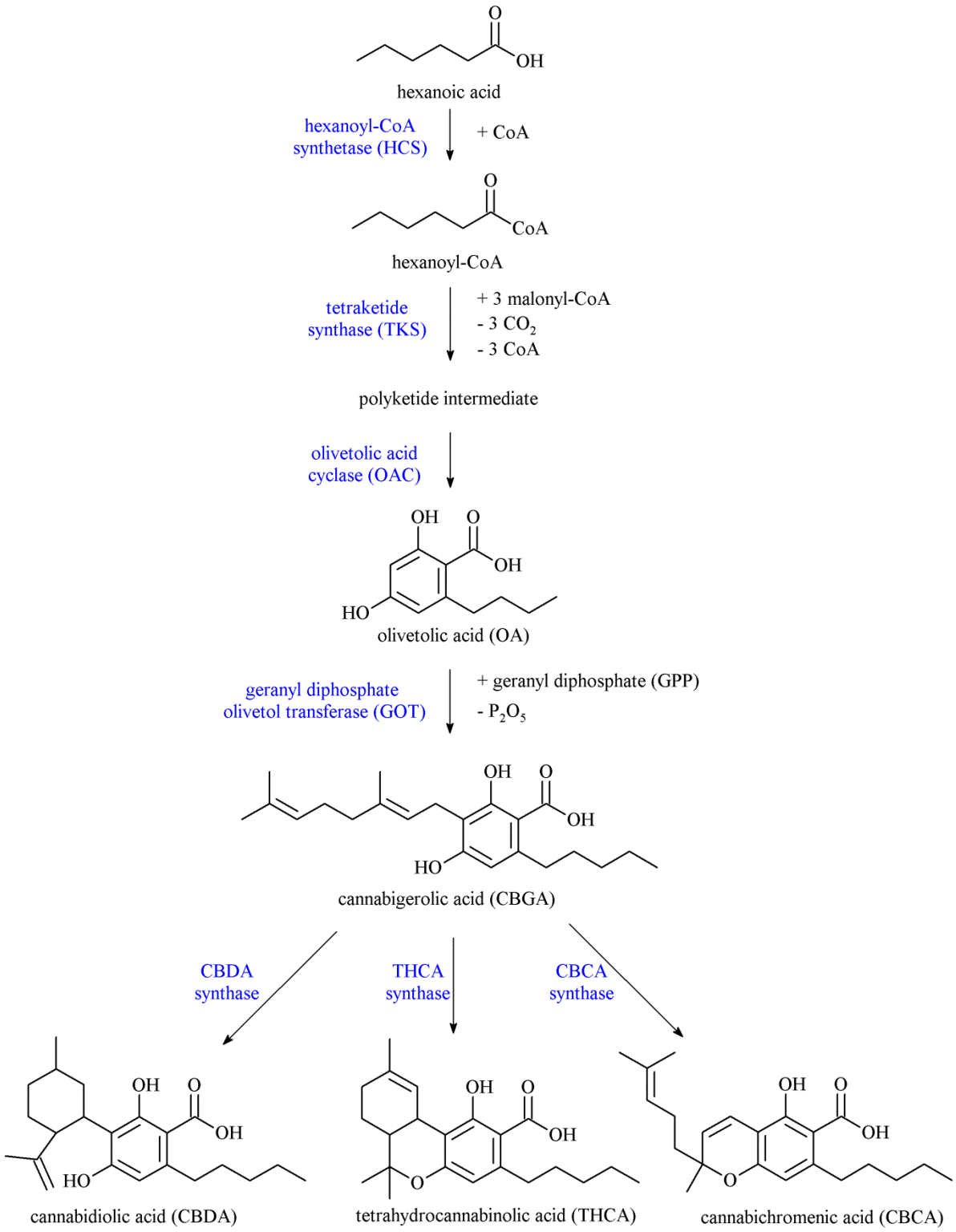


Figure 7.4 A proposed biosynthesis of cannabinoids.

The biosynthesis of cannabinoids is preceded by the biosynthesis of OA by TKS and OAC.

7.6 DABB Cyclases May Play a Large Role in Plants

Given their abundance and conservation, OAC-like DABB proteins are possibly involved in a variety of metabolic processes in plants other than cannabis. Although the occurrence of OA is rare, other stilbene carboxylic acids in plants have been detected and are known to occur, such as hydrangeic acid and lunularic acid in *Hydrangea macrophylla* (Zhang et al., 2009). Although one study has attributed STCS activity as being responsible for the production of stilbene carboxylates in *Hydrangea*, the lack of resolution and detection suggest another factor, or more specifically another protein with a similar function as OAC, is required for the biosynthesis of stilbene carboxylates (Eckermann et al., 2003). Similarly, the *in vitro* biosynthesis of derailed pyrones instead of anticipated end products, such as in the case of anthrone biosynthesis in *Aloe arborescens*, may require an OAC-like cyclase to fold large polyketide into their anticipated form (Mizuuchi et al., 2009). Furthermore, the associations of OAC-like proteins with plant stress and defense responses have been noted with regards to PtSP1 (Park et al., 2007; Zhu et al., 2008). Given the defense-inducing effects of jasmonic acid, a plant hormone associated with stress which is known to increase trichome production in plants, it is plausible that OAC-like proteins may be involved in the mediation of unfavorable conditions, such as salt and heat stress, in plants (Traw and Bergelson, 2003; van Schie et al., 2007). It is possible that OAC-like proteins are involved in the production of bioactive stilbene carboxylic acids in response to plant stresses in an attempt to produce antibiotic/antifungal/antiherbivore compounds in a more global context (Park et al., 2007).

7.7 Conclusion

This study set out to determine why OA was not being produced by a type III PKS implicitly involved in cannabinoid biosynthesis. The hypothesis that a putative cyclase was involved led to the discovery of OAC, a 12 kDa DABB protein belonging to the family of SRABB plant proteins of unknown function. The biochemical characterization of TKS and OAC has shown that, although TKS is capable of producing HTAL, PDAL, and olivetol, the inclusion of OAC during *in vitro* assays is required for the successful biosynthesis OA. Furthermore, TKS and OAC activities are physically independent. It is proposed that TKS generates a linear polyketide that is subsequently cyclized either spontaneously to a pyrone or olivetol as a result of enzyme product release, or the hydrolytic release of CoA, respectively. The fate of CoA-linked tetraketide intermediates is influenced in the presence of OAC, which is able to cyclize the intermediate to its proper resorcylic acid without a concurrent decarboxylation. In addition to proposing the mechanistic polyketide origins of OA and olivetol, this work also suggests a plausible mechanism for biosynthesis of lactones, alternatively known as derailment products, by type III PKS. The discovery of OAC also attributes the first known function to any protein belonging to the SRABB protein family. Lastly, this discovery resolves the enigmatic step of OA biosynthesis which is required during the biosynthesis of cannabinoids.

7.8 Future Directions

The discovery of OAC and its involvement in polyketide biosynthesis will undoubtedly play an important role in the identification and characterization of other OAC-like proteins in the future. The possibililty that other plant ‘type II’ PKCs are involved in type III PKS product cyclization is itself a novel finding that should generate a fresh perspective regarding plant natural product biosynthesis. It will be interesting to see if other plant OAC-like proteins are involved in

polyketide cyclizations. Regardless, OAC itself is an interesting protein, being the first plant polyketide cyclase that has ever been identified. Much more work will be required to clarify its mechanisms, which will necessitate more structure/function studies that will hopefully target and substitute other amino acids putatively involved in OAC activity. Also, the structural characterization of OAC should be conducted, either by NMR or by X-ray crystallography, to determine with highest certainty the positions and orientation of OAC's structural elements.

It will also be interesting to see if OAC is capable of participating in cyclization reactions not related to cannabinoid biosynthesis. For example, the effect that OAC has when it is assayed with other interesting plant PKSs, such as *Aloe arborescens* octaketide synthase, which accumulates SEK4 and SEK4b instead of chrysophanol anthrone, would be very interesting to see (Figure 2.7). Similarly, the role that OAC could have during other cyclization reactions that do not occur in plants, but in fungi and streptomycetes, will also be interesting to perform, since this could generate a plethora of new compounds, which could be of extreme medical benefit as potentially new antibiotics or anti-tumor compounds. It will be interesting to see how the role of OAC plays out in the future.

8. REFERENCES

Abe,I., Utsumi,Y., Oguro,S., Morita,H., Sano,Y., and Noguchi,H. (2005). A plant type III polyketide synthase that produces pentaketide chromone. J Am Chem Soc *127*, 1362-1363.

Abel,E.L. (1980). Marihuana, the first twelve thousand years. (Michigan: Plenum Press).

Adams,M.A. and Jia,Z. (2005). Structural and biochemical evidence for an enzymatic quinone redox cycle in *Escherichia coli*: identification of a novel quinol monooxygenase. J Biol Chem *280*, 8358-8363.

Adams,R. (1940). Marihuana. Science *92*, 115-119.

Adams,R., Hunt,M., and Clark,J.H. (1940a). Structure of cannabidiol, a product isolated from the marihuana extract of Minnesota Wild Hemp. J Am Chem Soc *62*, 196-200.

Adams,R., Pease,D.C., Cain,C.K., and Clark,J.H. (1940b). Structure of cannabidiol. VI. Isomerization of cannabidiol to tetrahydrocannabinol, a physiologically active product. Conversion of cannabidiol to cannabinol. J Am Chem Soc *62*, 2405-2408.

Ainsworth,C. (2000). Boys and girls come out to play: the molecular biology of dioecious plants. Ann Bot London *86*, 211-221.

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W., and Lipman,D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Amar,MB. (2006). Cannabinoids in medicine: A review of their therapeutic potential. J Ethnopharmacol *105*, 1-25.

Ames,B.D., Korman,T.P., Zhang,W., Smith,P., Vu,T., Tang,Y., and Tsai,S.C. (2008). Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides. Proc Natl Acad Sci USA *105*, 5349-5354.

Ames,B.D., Lee,M.Y., Moody,C., Zhang,W., Tang,Y., and Tsai,S.C. (2011). Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway. Biochemistry *50*, 8392-8406.

Armaleo,D., Sun,X., and Culberson,C. (2011). Insights from the first putative biosynthetic gene cluster for a lichen depside and depsidone. Mycologia *103*, 741-754.

Arnold,K., Bordoli,L., Kopp,J., and Schwede,T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics *22*, 195-201.

Austin,M.B., Bowman,M.E., Ferrer,J.L., Schröder,J., and Noel,J.P. (2004). An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. Chem Biol *11*, 1179-1194.

Austin,M.B. and Noel,J.P. (2003). The chalcone synthase superfamily of type III polyketide synthases. Nat Prod Rep *20*, 79-110.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A., and Struhl,K. (1994). Current protocols in molecular biology. (New York: Greene Publishing Associates and Wiley-Interscience).

Baerson,S.R., Schröder,J., Cook,D., Rimando,A.M., Pan,Z., Dayan,F.E., Noonan,B.P., and Duke,S.O. (2010). Alkylresorcinol biosynthesis in plants: new insights from an ancient enzyme family? Plant Signal Behav *5*, 1286-1289.

Bateman,A., Coin,L., Durbin,R., Finn,R.D., Hollich,V., Griffiths-Jones,S., Khanna,A., Marshall,M., Moxon,S., Sonnhammer,E.L., Studholme,D.J., Yeats,C., and Eddy,S.R. (2004). The Pfam protein families database. Nucleic Acids Res *32*, D138-D141.

Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., Rapp,B.A., and Wheeler,D.L. (2000). GenBank. Nucleic Acids Res *28*, 15-18.

Bentley,R. and Zwitkowits,P.M. (1967). Biosynthesis of tropolones in *Penicillium stipitatum*. VII. The formation of polyketide lactones and other nontropolone compounds as a result of ethionine inhibition. J Am Chem Soc *89*, 676-680.

Bingman,C.A., Johnson,K.A., Peterson,F.C., Frederick,R.O., Zhao,Q., Thao,S., Fox,B.G., Volkman,B.F., Jeon,W.B., Smith,D.W., Newman,C.S., Ulrich,E.L., Hegeman,A., Sussman,M.R., Markley,J.L., and Phillips,G.N., Jr. (2004). Crystal structure of the protein from gene At3g17210 of *Arabidopsis thaliana*. Proteins *57*, 218-220.

Breithaupt,C., Kurzbauer,R., Lilie,H., Schaller,A., Strassner,J., Huber,R., Macheroux,P., and Clausen,T. (2006). Crystal structure of 12-oxophytodienoate reductase 3 from tomato: self-inhibition by dimerization. Proc Natl Acad Sci USA *103*, 14337-14342.

Brown,I., Cascio,M.G., Rotondo,D., Pertwee,R.G., Heys,S.D., and Wahle,K.W. (2012). Cannabinoids and omega-3/6 endocannabinoids as cell death and anticancer modulators. Prog Lipid Res *52*, 80-109.

Bruce,N.C., Cain,R.B., Pieper,D.H., and Engesser,K.H. (1989). Purification and characterization of 4-methylmuconolactone methyl-isomerase, a novel enzyme of the modified 3-oxoadipate pathway in nocardioform actinomycetes. Biochem J *262*, 303-312.

Cahn,R.S. (1932). *Cannabis indica* resin.Part III.The constitution of cannabinol. J Chem Soc 1342-1353.

Callaway,J.C. (2004). Hempseed as a nutritional resource: An overview. Euphytica *140*, 65-72.

Camara,B., Bielecki,P., Kaminski,F., dos,S., V, Plumeier,I., Nikodem,P., and Pieper,D.H. (2007). A gene cluster involved in degradation of substituted salicylates via ortho cleavage in *Pseudomonas* sp. strain MT1 encodes enzymes specifically adapted for transformation of 4-methylcatechol and 3-methylmuconate. J Bacteriol *189*, 1664-1674.

Castoe,T.A., Stephens,T., Noonan,B.P., and Calestani,C. (2007). A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs. Gene *392*, 47-58.

Childs,K.L., Hamilton,J.P., Zhu,W., Ly,E., Cheung,F., Wu,H., Rabinowicz,P.D., Town,C.D., Buell,C.R., and Chan,A.P. (2007). The TIGR Plant Transcript Assemblies database. Nucleic Acids Res *35*, D846-D851.

Chung,J.Y., Fujii,I., Harada,S., Sankawa,U., and Ebizuka,Y. (2002). Expression, purification, and characterization of AknX anthrone oxygenase, which is involved in aklavinone biosynthesis in *Streptomyces galilaeus*. J Bacteriol *184*, 6115-6122.

Conte,M.G., Gaillard,S., Lanau,N., Rouard,M., and Perin,C. (2008). GreenPhylDB: a database for plant comparative genomics. Nucleic Acids Res *36*, D991-D998.

Crosby,J., Treadwell,T., Hammerton,M., Vasilakis,K., Crump,M.P., Williams,D.S., and Mann,S. (2012). Stabilization and enhanced reactivity of actinorhodin polyketide synthase minimal complex in polymer-nucleotide coacervate droplets. Chem Commun (Camb) *48*, 11832-11834.

Czichi,U. and Kindl,H. (1977). Phenylalanine ammonia lyase and cinnamic acid hydroxylases as assembled consecutive enzymes on microsomal membranes of cucumber cotyledons: cooperation and subcellular distribution. Planta *134*, 133-143.

Dayanandan,P. and Kaufman,P.B. (1976). Trichomes of *Cannabis sativa* L. (Cannabaceae). Am J Bot *63*, 578-591.

de Hoffmann,E. (1996). Tandem mass spectrometry: a primer. J Mass Sprectrom *31*, 129-137.

de Jesus,A.E., Hull,W.E., Steyn,P.S., van Heerden,F.R., and Vleggaar,R. (1982). Biosynthesis of viridicatumtoxin, a mycotoxin from Penicilium expansum. J Chem Soc Chem Comm *0*, 902-904.

Debnam,P.M., Shearer,G., Blackwood,L., and Kohl,D.H. (1997). Evidence for channeling of intermediates in the oxidative pentose phosphate pathway by soybean and pea nodule extracts, yeast extracts, and purified yeast enzymes. Eur J Biochem *246*, 283-290.

Devane,W.A., Dysarz,F.A., III, Johnson,M.R., Melvin,L.S., and Howlett,A.C. (1988). Determination and characterization of a cannabinoid receptor in rat brain. Mol Pharmacol *34*, 605-613.

Devane,W.A., Hanus,L., Breuer,A., Pertwee,R.G., Stevenson,L.A., Griffin,G., Gibson,D., Mandelbaum,A., Etinger,A., and Mechoulam,R. (1992). Isolation and structure of a brain constituent that binds to the cannabinoid receptor. Science *258*, 1946-1949.

Dgany,O., Gonzalez,A., Sofer,O., Wang,W., Zolotnitsky,G., Wolf,A., Shoham,Y., Altman,A., Wolf,S.G., Shoseyov,O., and Almog,O. (2004). The structural basis of the thermostability of SP1, a novel plant (*Populus tremula*) boiling stable protein. J Biol Chem *279*, 51516-51523.

Dobritsa,A.A., Lei,Z., Nishikawa,S.i., Urbanczyk-Wochniak,E., Huhman,D.V., Preuss,D., and Sumner,L.W. (2010). LAP5 and LAP6 encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in Arabidopsis. Plant Physiol *153*, 937-955.

Earley,K.W., Haag,J.R., Pontes,O., Opper,K., Juehne,T., Song,K., and Pikaard,C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. Plant J *45*, 616-629.

Eckermann,C., Schröder,G., Eckermann,S., Strack,D., Schmidt,J., Schneider,B., and Schröder,J. (2003). Stilbenecarboxylate biosynthesis: a new function in the family of chalcone synthase-related proteins. Phytochemistry *62*, 271-286.

Eckermann,S., Schröder,G., Schmidt,J., Strack,D., Edrada,R.A., Helariutta,Y., Elomaa,P., Kotilainen,M., Kilpelainen,I., Proksch,P., Teeri,T.H., and Schröder,J. (1998). New pathway to polyketides in plants. Nature *396*, 387-390.

Elix,J.A. (1974). Synthesis of para-olivetol depsides. Aust J Chem *27*, 1767-1779.

Elix,J.A. (1994). New depsides from the lichen *Lecanora planaica*. Aust J Chem *47*, 1199-1203.

Fellermeier,M., Eisenreich,W., Bacher,A., and Zenk,M.H. (2001). Biosynthesis of cannabinoids. Incorporation experiments with (13)C-labeled glucoses. Eur J Biochem *268*, 1596-1604.

Fellermeier,M. and Zenk,M.H. (1998). Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. FEBS Lett *427*, 283-285.

Felsenstein,J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution *39*, 783-791.

Fesner,W.D., Schneider,A., Held,H., Sinerius,G., Walter,C., Hixon,M., Schloss,J.V. (1996). The Mechanism of Class II, Metal-Dependent Aldolases. Angewandte Chemie International Edition in English *35*, 2219-2221

Fetterman,P.S., Keith,E.S., Waller,C.W., Guerrero,O., Doorenbos,N.J., and Quimby,M.W. (1971). Mississippi-grown *Cannabis sativa* L preliminary observation on chemical definition of phenotype and variations in tetrahydrocannabinol content versus age, sex, and plant part. J Pharm Sci *60*, 1246-1249.

Fleming,M.P. and Clarke,R.C. (1998). Physical evidence for the antiquity of *Cannabis sativa* L. Journal of Industrial Hemp *5*, 80-92.

Flores-Sanchez,I.J., Linthorst,H.J., and Verpoorte,R. (2010). In silicio expression analysis of PKS genes isolated from *Cannabis sativa* L. Genet Mol Biol *33*, 703-713.

Flores-Sanchez,I.J. and Verpoorte,R. (2008a). PKS activities and biosynthesis of cannabinoids and flavonoids in *Cannabis sativa* L plants. Plant Cell Physiol *49*, 1767-1782.

Flores-Sanchez,I.J. and Verpoorte,R. (2008b). Secondary metabolism in Cannabis. Phytochem Rev. *7*, 615-639.

Fridman,E. and Pichersky,E. (2005). Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. Curr Opin Plant Biol *8*, 242-248.

Funa,N., Awakawa,T., and Horinouchi,S. (2007). Pentaketide resorcylic acid synthesis by type III polyketide synthase from *Neurospora crassa*. J Biol Chem *282*, 14476-14481.

Gagne,S.J., Stout,J.M., Liu,E., Boubakir,Z., Clark, S.M., and Page,J.E. (2012) Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. Proc Natl Acad Sci USA *109*, 12811-12816

Gang,D.R., Wang,J., Dudareva,N., Nam,K.H., Simon,J.E., Lewinsohn,E., and Pichersky,E. (2001). An investigation of the storage and biosynthesis of phenylpropenes in sweet basil. Plant Physiol *125*, 539-555.

Gaoni,Y. and Mechoulam,R. (1964). Isolation, structure, and partial synthesis of an active constituent of hashish. J Am Chem Soc *86*, 1646-1647.

Gaoni,Y. and Mechoulam,R. (1971). The isolation and structure of delta-1-tetrahydrocannabinol and other neutral cannabinoids from hashish. J Am Chem Soc *93*, 217-224.

Gensheimer,M. and Mushegian,A. (2004). Chalcone isomerase family and fold: no longer unique to plants. Protein Sci *13*, 540-544.

Gershenzon,J., McCaskill,D., Rajaonarivony,J.I., Mihaliak,C., Karp,F., and Croteau,R. (1992). Isolation of secretory cells from plant glandular trichomes and their use in biosynthetic studies of monoterpenes and other gland products. Anal Biochem *200*, 130-138.

Gijsen,H.J., Qiao,L., Fitz,W., and Wong,C.H. (1996). Recent advances in the chemoenzymatic synthesis of carbohydrates and carbohydrate mimetics. Chem Rev *96*, 443-474.

Glish,G.L. and Vachet,R.W. (2003). The basics of mass spectrometry in the twenty-first century. Nat Rev Drug Discov *2*, 140-150.

Goblirsch,B.R., Streit,B.R., Dubois,J.L., and Wilmot,C.M. (2010). Structural features promoting dioxygen production by *Dechloromonas aromatica* chlorite dismutase. J Biol Inorg Chem *15*, 879-888.

Gomori,G. (1955). Preparation of buffers for use in enzyme studies. Methods Enzymol *1*, 138-146.

Goodstein,D.M., Shu,S., Howson,R., Neupane,R., Hayes,R.D., Fazo,J., Mitros,T., Dirks,W., Hellsten,U., Putnam,N., and Rokhsar,D.S. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res *40*, D1178-D1186.

Gorham,J. (1995). The biochemistry of the stilbenoids. (London: Chapman and Hall).

Griesebach H, Fritsch H. (1975). Biosynthesis of cyanidin in cell cultures of *Haplopappus gracilis*. Phytochemistry *14***,** 2437-2442.

Guilford,H., Scott,A.I., Skingle,D., and Yalpani,M. (1968). The synthesis of tetra-acetic acid lactone and a model for the biosynthesis of 6-methylsalicyclic acid. Chem Commun (London) 1127-1128.

Hadfield,A.T., Limpkin,C., Teartasin,W., Simpson,T.J., Crosby,J., and Crump,M.P. (2004). The crystal structure of the ActIII actinorhodin polyketide reductase: proposed mechanism for ACP and polyketide binding. Structure *12*, 1865-1875.

Hammond,C.T. and Mahlberg,P.G. (1973). Morphology of glandular hairs of *Cannabis sativa* from scanning electron microscopy. Am J Bot *60*, 524-528.

Hammond,C.T. and Mahlberg,P.G. (1978). Ultrastructural development of capitate glandular hairs of *Cannabis Sativa* L. (Cannabaceae). Am J Bot *65*, 140-151.

Harris,T.M. and Carney,R.L. (1966). Biogenetically modeled synthesis of β-resorcylic acids. J Am Chem Soc *88*, 2053-2054.

Harris,T.M. and Carney,R.L. (1967). Synthesis of 3,5,7-triketo acids and esters and their cyclizations to resorcinol and phloroglucinol derivatives. Models of biosynthesis of phenolic compounds. J Am Chem Soc *89*, 6734-6741.

Harvey,D.J. (1976). Characterization of the butyl homologues of delta1-tetrahydrocannabinol, cannabinol and cannabidiol in samples of cannabis by combined gas chromatography and mass spectrometry. J Pharm Pharmacol *28*, 280-285.

Hassani,I. and Razzoul,H. (2005). Olivetol: constituent of lichen *Evernia Prunastri* Ach. Physical and Chemical News *26*, 98-103.

Hathway,D.E. and Seakin,J.W. (1959). Hydroxystilbenes of *Eucalyptus wandoo*. Biochem J *72*, 369-374.

Hill,A.M. and Staunton,J. (2010). 1.10 - Type I modular PKS. In Comprehensive Natural Products II, M.Editors-in-Chief:Â Â Lew and B.Hung-Wen, eds. (Oxford: Elsevier), pp. 385-452.

Horper,W. and Marner,F.J. (1995). Phenols and quinones from leaves of *Primula obconica*. Nat Prod Lett *6*, 163-170.

Horper,W. and Marner,F.J. (1996). Biosynthesis of primin and micodnidin and its derivatives. Phytochemistry *41*, 451-456.

Hrazdina,G., Wagner,G.J., Siegelman,H.W. (1978). Subcellular localization of enzymes of anthocyanin biosynthesis in protoplasts. Phytochemistry *17*, 53-56.

Iijima,Y., Davidovich-Rikanati,R., Fridman,E., Gang,D.R., Bar,E., Lewinsohn,E., and Pichersky,E. (2004). The biochemical and molecular basis for the divergent patterns in the biosynthesis of terpenes and phenylpropenes in the peltate glands of three cultivars of basil. Plant Physiol *136*, 3724-3736.

Intrieri,M.C. and Buiatti,M. (2001). The horizontal transfer of *Agrobacterium rhizogenes* genes and the evolution of the genus *Nicotiana*. Mol Phylogenet Evol *20*, 100-110.

Isbell,H. (1973). Research on Cannabis (marijuana). UN Bulletin on Narcotics *25*, 37-48.

Jenke-Kodama,H., Sandmann,A., Muller,R., and Dittmann,E. (2005). Evolutionary implications of bacterial polyketide synthases. Mol Biol Evol *22*, 2027-2039.

Jez,J.M., Austin,M.B., Ferrer,J., Bowman,M.E., Schröder,J., and Noel,J.P. (2000). Structural control of polyketide formation in plant-specific polyketide synthases. Chem Biol *7*, 919-930.

Jez,J.M., Ferrer,J.L., Bowman,M.E., Austin,M.B., Schröder,J., Dixon,R.A., and Noel,J.P. (2001). Structure and mechanism of chalcone synthase-like polyketide synthases. J Ind Microbiol Biotechnol *27*, 393-398.

Jiang,H.E., Li,X., Zhao,Y.X., Ferguson,D.K., Hueber,F., Bera,S., Wang,Y.F., Zhao,L.C., Liu,C.J., and Li,C.S. (2006). A new insight into *Cannabis sativa* (Cannabaceae) utilization from 2500-year-old Yanghai Tombs, Xinjiang, China. J Ethnopharmacol *108*, 414-422.

Jin,W. and Zjawiony,J.K. (2006). 5-alkylresorcinols from *Merulius incarnatus*. J Nat Prod *69*, 704-706.

Jirschitzka,J., Mattern,D.J., Gershenzon,J., and D'Auria,J.C. (2013). Learning from nature: new approaches to the metabolic engineering of plant defense pathways. Curr Opin Biotechnol *24*, 320-328.

Jones,D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol *292*, 195-202.

Jones,D.T., Taylor,W.R., and Thornton,J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci *8*, 275-282.

Jones, M. Isolation of plasmid DNA from yeast using the QIAprep® Spin Miniprep Kit. 2000. (PR04.doc Oct-01), QIAGEN.

Jones,T.H., Brunner,S.R., Edwards,A.A., Davidson,D.W., and Snelling,R.R. (2005). 6-alkylsalicylic acids and 6-alkylresorcylic acids from ants in the genus *Crematogaster* from Brunei. J Chem Ecol *31*, 407-417.

Karppinen,K. and Hohtola,A. (2008). Molecular cloning and tissue-specific expression of two cDNAs encoding polyketide synthases from *Hypericum perforatum*. J Plant Physiol *165*, 1079-1086.

Keatinge-Clay,A.T. (2012). The structures of type I polyketide synthases. Nat Prod Rep *29*, 1050-1073.

Keene,C.K. and Wagner,G.J. (1985). Direct demonstration of duvatrienediol biosynthesis in glandular heads of tobacco trichomes. Plant Physiol *79*, 1026-1032.

Kendrew,S.G., Hopwood,D.A., and Marsh,E.N. (1997). Identification of a monooxygenase from *Streptomyces coelicolor* A3(2) involved in biosynthesis of actinorhodin: purification and characterization of the recombinant enzyme. J Bacteriol *179*, 4305-4310.

Kim,E.S. and Mahlberg,P.G. (1997). Plastid development in disc cells of glandular trichomes of Cannabis (Cannabaceae). Mol Cells *7*, 352-359.

Kim,M.I., Kwon,S.J., and Dordick,J.S. (2009). In vitro precursor-directed synthesis of polyketide analogues with coenzyme a regeneration for the development of antiangiogenic agents. Org Lett *11*, 3806-3809.

Kimura,M. and Okamoto,K. (1970). Distribution of tetrahydrocannabinolic acid in fresh wild cannabis. Experientia *26*, 819-820.

Kozubek,A. and Tyman,J.H. (1999). Resorcinolic lipids, the natural non-isoprenoid phenolic amphiphiles and their biological activity. Chem Rev *99*, 1-26.

Lee,W.C., Reniere,M.L., Skaar,E.P., and Murphy,M.E. (2008). Ruffling of metalloporphyrins bound to IsdG and IsdI, two heme-degrading enzymes in *Staphylococcus aureus*. J Biol Chem *283*, 30957-30963.

Loewe,S. (1945). Marihuana activity of cannabinol. Science *14*, 615-616.

Ma,S.M., Zhan,J., Xie,X., Watanabe,K., Tang,Y., and Zhang,W. (2007). Redirecting the cyclization steps of fungal polyketide synthase. J Am Chem Soc *130*, 38-39.

Mandolino,G., Carboni,A., Bagatta,M., Moliterni,V.M., and Ranalli,P. (2002). Occurrence and frequency of putatively Y chromosome linked DNA markers in *Cannabis sativa* L. Euphytica *126*, 211-218.

Marchler-Bauer,A., Lu,S., Anderson,J.B., Chitsaz,F., Derbyshire,M.K., Weese-Scott,C., Fong,J.H., Geer,L.Y., Geer,R.C., Gonzales,N.R., Gwadz,M., Hurwitz,D.I., Jackson,J.D., Ke,Z., Lanczycki,C.J., Lu,F., Marchler,G.H., Mullokandov,M., Omelchenko,M.V., Robertson,C.L.,

Song,J.S., Thanki,N., Yamashita,R.A., Zhang,D., Zhang,N., Zheng,C., and Bryant,S.H. (2011). CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res *39*, D225-D229.

Marks,M.D., Tian,L., Wenger,J.P., Omburo,S.N., Soto-Fuentes,W., He,J., Gang,D.R., Weiblen,G.D., and Dixon,R.A. (2009). Identification of candidate genes affecting Δ9-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. J Exp Bot *60*, 3715-3726.

Martinez,M. (2011). Plant protein-coding gene families: emerging bioinformatics approaches. Trends Plant Sci *16*, 558-567.

Matousek,J., Vrba,L., Novak,P., Patzak,J., De,K.J., Skopek,J., Heyerick,A., Roldan-Ruiz,I., and De,K.D. (2005). Cloning and molecular analysis of the regulatory factor HlMyb1 in hop (*Humulus lupulus* L.) and the potential of hop to produce bioactive prenylated flavonoids. J Agric Food Chem *53*, 4793-4798.

Matsuzawa,M., Katsuyama,Y., Funa,N., and Horinouchi,S. (2010). Alkylresorcylic acid synthesis by type III polyketide synthases from rice *Oryza sativa*. Phytochemistry *71*, 1059-1067.

Mechoulam,R. (1988). Alkaloids in *Cannabis sativa* L. In The Alkaloids: Chemistry and Pharmacology, A.Brossi, ed. (San Diego: Academic Press), pp. 77-93.

Mechoulam,R., Ben-Shabat,S., Hanus,L., Ligumsky,M., Kaminski,N.E., Schatz,A.R., Gopher,A., Almog,S., Martin,B.R., and Compton,D.R. (1995). Identification of an endogenous 2-monoglyceride, present in canine gut, that binds to cannabinoid receptors. Biochem Pharmacol *50*, 83-90.

Mechoulam,R. and Hanus,L. (2000). A historical overview of chemical research on cannabinoids. Chem Phys Lipids *108*, 1-13.

Merkus,F.W. (1971). Cannabivarin and tetrahydrocannabivarin, two new constituents of hashish. Nature *232*, 579-580.

Mizuuchi,Y., Shi,S.P., Wanibuchi,K., Kojima,A., Morita,H., Noguchi,H., and Abe,I. (2009). Novel type III polyketide synthases from *Aloe arborescens*. FEBS lett *276*, 2391-2401.

Moliterni,V., Cattivelli,L., Ranalli,P., and Mandolino,G. (2004). The sexual differentiation of *Cannabis sativa* L.: A morphological and molecular study. Euphytica *140*, 95-106.

Morimoto,S., Komatsu,K., Taura,F., and Shoyama,Y. (1998). Purification and characterization of cannabichromenic acid synthase from *Cannabis sativa*. Phytochemistry *49*, 1525-1529.

Morimoto,S., Komatsu,K., Taura,F., and Shoyama,Y. (1997). Enzymological evidence for cannabichromenic acid biosynthesis. J Nat Prod *60*, 854-857.

Muggia,L. and Grube,M. (2010). Type III polyketide synthases in lichen mycobionts. Fungal Biol *114*, 379-385.

Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bradley,P., Bork,P., Bucher,P., Cerutti,L., Copley,R., Courcelle,E., Das,U., Durbin,R., Fleischmann,W., Gough,J., Haft,D., Harte,N., Hulo,N., Kahn,D., Kanapin,A., Krestyaninova,M., Lonsdale,D., Lopez,R., Letunic,I., Madera,M., Maslen,J., McDowall,J., Mitchell,A., Nikolskaya,A.N., Orchard,S., Pagni,M., Ponting,C.P., Quevillon,E., Selengut,J., Sigrist,C.J., Silventoinen,V., Studholme,D.J., Vaughan,R., and Wu,C.H. (2005). InterPro, progress and status in 2005. Nucleic Acids Res *33*, D201-D205.

Munro,S., Thomas,K.L., and Abu-Shaar,M. (1993). Molecular characterization of a peripheral receptor for cannabinoids. Nature *365*, 61-65.

Munson,A.E., Harris,L.S., Friedman,M.A., Dewey,W.L., and Carchman,R.A. (1975). Antineoplastic activity of cannabinoids. J Natl Cancer Inst *55*, 597-602.

Murzin,A.G., Brenner,S.E., Hubbard,T., and Chothia,C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol *247*, 536-540.

Nagel,J., Culley,L.K., Lu,Y., Liu,E., Matthews,P.D., Stevens,J.F., and Page,J.E. (2008). EST analysis of hop glandular trichomes identifies an O-methyltransferase that catalyzes the biosynthesis of xanthohumol. Plant Cell *20*, 186-200.

O'Shaughnessy,W.B. (1843). On the preparations of the Indian hemp, or gunjah. Prov Med J Retrosp Med Sci *5*, 363-369.

Orengo,C.A. and Thornton,J.M. (1993). Alpha plus beta folds revisited: some favoured motifs. Structure *1*, 105-120.

Paniego,N.B., Zuurbier,K.W., Fung,S.Y., van der,H.R., Scheffer,J.J., and Verpoorte,R. (1999). Phlorisovalerophenone synthase, a novel polyketide synthase from hop (*Humulus lupulus* L.) cones. Eur J Biochem *262*, 612-616.

Park,S.C., Lee,J.R., Shin,S.O., Park,Y., Lee,S.Y., and Hahm,K.S. (2007). Characterization of a heat-stable protein with antimicrobial activity from *Arabidopsis thaliana*. Biochem Biophys Res Commun *362*, 562-567.

Pelletier,M.K., Murrell,J.R., and Shirley,B.W. (1997). Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in Arabidopsis. Further evidence for differential regulation of "early" and "late" genes. Plant Physiol *113*, 1437-1445.

Pertwee,R.G., Howlett,A.C., Abood,M.E., Alexander,S.P., Di,M., V, Elphick,M.R., Greasley,P.J., Hansen,H.S., Kunos,G., Mackie,K., Mechoulam,R., and Ross,R.A. (2010). International Union of Basic and Clinical Pharmacology. LXXIX. Cannabinoid receptors and their ligands: beyond CB(1) and CB(2). Pharmacol Rev *62*, 588-631.

Petkovic,H., Thamchaipenet,A., Zhou,L.H., Hranueli,D., Raspor,P., Waterman,P.G., and Hunter,I.S. (1999). Disruption of an aromatase/cyclase from the oxytetracycline gene cluster of *Streptomyces rimosus* results in production of novel polyketides with shorter chain lengths. J Biol Chem *274*, 32829-32834.

Pieper,D.H., Stadler-Fritzsche,K., Knackmuss,H.J., Engesser,K.H., Bruce,N.C., and Cain,R.B. (1990). Purification and characterization of 4-methylmuconolactone methylisomerase, a novel enzyme of the modified 3-oxoadipate pathway in the gram-negative bacterium *Alcaligenes eutrophus* JMP 134. Biochem J *271*, 529-534.

Potter, D. J. P. The propagation, characterisation and optimisation of *Cannabis sativa* L. as a phytopharmaceutical. 2009. King's College London.

Prabhu,V., Chatson,K.B., Abrams,G.D., and King,J. (1996). 13C nuclear magnetic resonance detection of interactions of serine hydroxymethyltransferase with C1-tetrahydrofolate synthase and glycine decarboxylase complex activities in Arabidopsis. Plant Physiol *112*, 207-216.

Radauer,C., Lackner,P., and Breiteneder,H. (2008). The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. BMC Evol Biol *8*, 286.

Raharjo,T.J., Chang,W.T., Choi,Y.H., Peltenburg-Looman,A.M.G., and Verpoorte,R. (2004a). Olivetol as product of a polyketide synthase in *Cannabis sativa* L. Plant Sci *166*, 381-385.

Raharjo,T.J., Chang,W.T., Verberne,M.C., Peltenburg-Looman,A.M., Linthorst,H.J., and Verpoorte,R. (2004b). Cloning and over-expression of a cDNA encoding a polyketide synthase from *Cannabis sativa*. Plant Physiol Biochem *42*, 291-297.

Raty,K., Kantola,J., Hautala,A., Hakala,J., Ylihonko,K., and Mantsala,P. (2002). Cloning and characterization of *Streptomyces galilaeus* aclacinomycins polyketide synthase (PKS) cluster. Gene *293*, 115-122.

Russo,E.B. (2011). Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. Br J Pharmacol *163*, 1344-1364.

Russo,E.B., Jiang,H.E., Li,X., Sutton,A., Carboni,A., del,B.F., Mandolino,G., Potter,D.J., Zhao,Y.X., Bera,S., Zhang,Y.B., Lu,E.G., Ferguson,D.K., Hueber,F., Zhao,L.C., Liu,C.J., Wang,Y.F., and Li,C.S. (2008). Phytochemical and genetic analyses of ancient cannabis from Central Asia. J Exp Bot *59*, 4171-4182.

Saito,K., Yokoyama,H., Noji,M., and Murakoshi,I. (1995). Molecular cloning and characterization of a plant serine acetyltransferase playing a regulatory role in cysteine biosynthesis from watermelon. J Biol Chem *270*, 16321-16326.

Sallaud,C., Rontein,D., Onillon,S., Jabes,F., Duffe,P., Giacalone,C., Thoraval,S., Escoffier,C., Herbette,G., Leonhardt,N., Causse,M., and Tissier,A. (2009). A novel pathway for sesquiterpene biosynthesis from Z,Z-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. Plant Cell *21*, 301-317.

Schilmiller,A.L., Last,R.L., and Pichersky,E. (2008). Harnessing plant trichome biochemistry for the production of useful compounds. Plant J *54*, 702-711.

Schneider,G. (2005). Enzymes in the biosynthesis of aromatic polyketide antibiotics. Curr Opin Struct Biol *15*, 629-636.

Schröder,G., Brown,J.W., and Schröder,J. (1988). Molecular analysis of resveratrol synthase. cDNA, genomic clones and relationship with chalcone synthase. Eur J Biochem *172*, 161-169.

Schröder, J. Relationship tree of selected plant type III PKS. 2010. http://www.biologie.uni-freiburg.de, http://www.biologie.uni-freiburg.de.

Schröder, J. One, two, and three condensations. 2009. http://www.biologie.uni-freiburg.de/data/bio2/schroeder/Homepage_English.html, http://www.biologie.uni-freiburg.de.

Schultes,R.E., Klein,W.M., Plowman,T., and Lockwood,T.E. (1974). Cannabis: an example of taxonomic neglect. Botanical Museum Leaflets. [Harvard University]. Cambridge, MA *23* , 337-367.

Sciara,G., Kendrew,S.G., Miele,A.E., Marsh,N.G., Federici,L., Malatesta,F., Schimperna,G., Savino,C., and Vallone,B. (2003). The structure of ActVA-Orf6, a novel type of monooxygenase involved in actinorhodin biosynthesis. EMBO J *22*, 205-215.

Seigler,D.S. (1998). Plant secondary metabolism. (Boston: Kluwer Academic).

Shen,B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr Opin Chem Biol *7*, 285-295.

Shen,B. and Hutchinson,C.R. (1996). Deciphering the mechanism for the assembly of aromatic polyketides by a bacterial polyketide synthase. Proc Natl Acad Sci USA *93*, 6600-6604.

Shen,B. and Hutchinson,C.R. (1993b). Tetracenomycin F2 cyclase: intramolecular aldol condensation in the biosynthesis of tetracenomycin C in *Streptomyces glaucescens*. Biochemistry *32*, 11149-11154.

Shen,B. and Hutchinson,C.R. (1993a). Tetracenomycin F1 monooxygenase: oxidation of a naphthacenone to a naphthacenequinone in the biosynthesis of tetracenomycin C in *Streptomyces glaucescens*. Biochemistry *32*, 6656-6663.

Shoyama,Y., Yagi,M., Nishioka,I., and Yamauchi,T. (1975). Biosynthesis of cannabinoid acids. Phytochemistry *14*, 2189-2192.

Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M., Soding,J., Thompson,J., and Higgins,D. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol *7*.

Silverman,R.B. (2002). Aldol and Claisen reactions and retroreactions. In The Organic Chemistry of Enzyme-Catalyzed Reactions, 2, ed. (Waltham. Mass, Academic Press), pp. 453-478.

Simonsen,J.L. and Todd,A.R. (1942). Cannabis indica. Part X. The essential oil from Egyptian hashish. J Chem Soc *1* , 188-191.

Small,E. and Beckstead,H.D. (1973). Common cannabinoid phenotypes in 350 stocks of Cannabis. Lloydia *36*, 144-165.

Soltis,D.E., Smith,S.A., Cellinese,N., Wurdack,K.J., Tank,D.C., Brockington,S.F., Refulio-Rodriguez,N.F., Walker,J.B., Moore,M.J., Carlsward,B.S., Bell,C.D., Latvis,M., Crawley,S., Black,C., Diouf,D., Xi,Z., Rushworth,C.A., Gitzendanner,M.A., Sytsma,K.J., Qiu,Y.L., Hilu,K.W., Davis,C.C., Sanderson,M.J., Beaman,R.S., Olmstead,R.G., Judd,W.S., Donoghue,M.J., and Soltis,P.S. (2011). Angiosperm phylogeny: 17 genes, 640 taxa. Am J Bot *98*, 704-730.

Song,B.H., Wang,X.Q., Li,F.Z., and Hong,D.Y. (2001). Further evidence for paraphyly of the Celtidaceae from the chloroplast gene *mat* K. Plant Syst Evol *228*, 107-115.

Sparkes,I.A., Runions,J., Kearns,A., and Hawes,C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc *1*, 2019-2025.

Stafford,H.A. (1974). The metabolism of aromatic compounds. Annu Rev Plant Physiol *25*, 459-486.

Staunton,J. and Weissman,K.J. (2001). Polyketide biosynthesis: a millennium review. Nat Prod Rep *18*, 380-416.

Stearn,W.T. (1970). The Cannabis plant: botanical characteristics. In The Botany and Chemistry of *Cannabis*., C.R.B.Joyce and S.H.Curry, eds. (London: J. & A. Churchill), pp. 1-10.

Stout,J.M., Boubakir,Z., Ambrose,S.J., Purves,R.W., and Page,J.E. (2012). The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in *Cannabis sativa* trichomes. Plant J *71*, 353-365.

Sugano,Y., Muramatsu,R., Ichiyanagi,A., Sato,T., and Shoda,M. (2007). DyP, a unique dye-decolorizing peroxidase, represents a novel heme peroxidase family: ASP171 replaces the distal histidine of classical peroxidases. J Biol Chem *282*, 36652-36658.

Sultana,A., Kallio,P., Jansson,A., Wang,J.S., Niemi,J., Mantsala,P., and Schneider,G. (2004). Structure of the polyketide cyclase SnoaL reveals a novel mechanism for enzymatic aldol condensation. EMBO J *23*, 1911-1921.

Suss,K.H., Arkona,C., Manteuffel,R., and Adler,K. (1993). Calvin cycle multienzyme complexes are bound to chloroplast thylakoid membranes of higher plants in situ. Proc Natl Acad Sci USA *90*, 5514-5518.

Sytsma,K.J., Morawetz,J., Pires,J.C., Nepokroeff,M., Conti,E., Zjhra,M., Hall,J.C., and Chase,M.W. (2002). Urticalean rosids: circumscription, rosid ancestry, and phylogenetics based on rbcL, trnL-F, and ndhF sequences. Am J Bot *89*, 1531-1546.

Taguchi,C., Taura,F., Tamada,T., Shoyama,Y., Shoyama,Y., Tanaka,H., Kuroki,R., and Morimoto,S. (2008). Crystallization and preliminary X-ray diffraction studies of polyketide synthase-1 (PKS-1) from *Cannabis sativa*. Acta Crystallogr Sect F Struct Biol Cryst Commun *64*, 217-220.

Takaoka,M. (1939). Resveratrol, a new phenolic compound, from *Veratrum grandiflorum*. Nippon Kagaku Kaishi *60*, 1090-1100.

Tamura,K., Peterson,D., Peterson,N., Stecher,G., Nei,M., and Kumar,S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol *28*, 2731-2739.

Taura,F., Morimoto,S., and Shoyama,Y. (1995). First direct evidence for the mechanism of delta-1-tetrahydrocannabinolic acid biosynthesis. J Am Chem Soc *117*, 9766-9767.

Taura,F., Morimoto,S., and Shoyama,Y. (1996). Purification and characterization of cannabidiolic-acid synthase from *Cannabis sativa* L.. Biochemical analysis of a novel enzyme that catalyzes the oxidocyclization of cannabigerolic acid to cannabidiolic acid. J Biol Chem *271*, 17411-17416.

Taura,F., Sirikantaramas,S., Shoyama,Y., Shoyama,Y., and Morimoto,S. (2007). Phytocannabinoids in *Cannabis sativa*: recent studies on biosynthetic enzymes. Chem Biodivers *4*, 1649-1663.

Taura,F., Tanaka,S., Taguchi,C., Fukamizu,T., Tanaka,H., Shoyama,Y., and Morimoto,S. (2009). Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. FEBS Lett *583*, 2061-2066.

Thomas,R. (2001). A biosynthetic classification of fungal and streptomycete fused-ring aromatic polyketides. ChemBioChem *2*, 612-627.

Thomas,R. and Williams,D.J. (1983). Oxytetracycline biosynthesis: mode of incorporation of [1-13C]- and [1,2-13C2]-acetate. J Chem Soc Chem Comm 128-130.

Thompson,T.B., Katayama,K., Watanabe,K., Hutchinson,C.R., and Rayment,I. (2004). Structural and functional analysis of tetracenomycin F2 cyclase from *Streptomyces glaucescens*. A type II polyketide cyclase. J Biol Chem *279*, 37956-37963.

Toffoli,F., Avico,U., and Signoretti-Ciranni,E. (1968). Methods of distinguishing biologically active cannabis and fibre cannabis. Bull Narc *1*, 55-59.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature *485*, 635-641.

Touwn,M. (1981). The religious and medical uses of Cannabis in China, India and Tibet. J Psychoactive Drugs *13*, 23-24.

Traw,M.B. and Bergelson,J. (2003). Interactive effects of jasmonic acid, salicylic acid, and gibberellin on induction of trichomes in Arabidopsis. Plant Physiol *133*, 1367-1375.

Turner,C.E., Elsohly,M.A., and Boeren,E.G. (1980). Constituents of *Cannabis sativa* L. XVII. A review of the natural constituents. J Nat Prod *43*, 169-234.

Ugwu,S.O. and Apte,S.P. (2004). The effect of buffers on protein conformational stability. Pharm Technol *28*, 86.

Van Bakel H., Stout,J.M., Cote,A.G., Tallon,C.M., Sharpe,A.G., Hughes,T.R., and Page,J.E. (2011). The draft genome and transcriptome of *Cannabis sativa*. Genome Biol *12*, R102.

van Schie,C.C., Haring,M.A., and Schuurink,R.C. (2007). Tomato linalool synthase is induced in trichomes by jasmonic acid. Plant Mol Biol *64*, 251-263.

Verpoorte,R., Alfermann,A.W., and Johnson,T.S. (2007). Applications of Plant Metabolic Engineering. Springer).

Vreet,T.B., Breimer,D., VanGinneken,C.A.M., and VanRossum,J.M. (1972). Identification in hashish of tetrahydrocannabinol, cannabidiol and cannabinol analogues with a methyl side-chain. J Pharm Sci *24*, 7-12.

Wagner,G.J. (1991). Secreting glandular trichomes: more than just hairs. Plant Physiol *96*, 675-679.

Wagner,G.J., Hrazdina,G. (1984). Endoplasmic Reticulum as a Site of Phenylpropanoid and Flavonoid Metabolism in *Hippeastrum*. Plant Physiol *74*, 901-906.

Wakimoto,T., Morita,H., and Abe,I. (2012). Engineering of plant type III polyketide synthases. Methods Enzymol *515*, 337-358.

Walter,S. and Buchner,J. (2002). Molecular chaperones--cellular machines for protein folding. Angew Chem Int Ed Engl *41*, 1098-1113.

Ward,J.J., Sodhi,J.S., Guffin,L.J., Buxton,B.F., and Jones,D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol *337*, 635-645.

Weigel,D. and Glazebrook,J. (2006). Transformation of agrobacterium using electroporation. CSH Protoc *2006*.

Weissman,K.J. (2009). Introduction to polyketide biosynthesis. In Methods Enzymol, David A.Hopwood, ed. Academic Press), pp. 3-16.

Winkel-Shirley,B. (1999). Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. Physiol Plant *107* , 142-149.

Wollner,H.J., Matchett,J.R., Levine,J., and Loewe,S. (1942). Isolation of a physiologically active tetrahydrocannabinol from *Cannabis sativa* resin. J Am Chem Soc *64*, 26-29.

Wood,B.T., Spivey,N.W.T., and Easterfield,T.H. (1896). Charas. The resin of Indian hemp. J Chem Soc Trans *69*, 539-546.

Wu,R., Skaar,E.P., Zhang,R., Joachimiak,G., Gornicki,P., Schneewind,O., and Joachimiak,A. (2005). *Staphylococcus aureus* IsdG and IsdI, heme-degrading enzymes with structural similarity to monooxygenases. J Biol Chem *280*, 2840-2846.

Yamashita,Y., Matsunami,K., Otsuka,H., Shinzato,T., and Takeda,Y. (2008). Grevillosides A-F: glucosides of 5-alkylresorcinol derivatives from leaves of *Grevillea robusta*. Phytochemistry *69*, 2749-2752.

Yamashita,Y., Matsunami,K., Otsuka,H., Shinzato,T., and Takeda,Y. (2010). 5-Alkylresorcinol glucosides from the leaves of *Grevillea robusta* Allan Cunningham. J Nat Med *64*, 474-477.

Yamauchi,T., Shoyama,Y., Aramaki,H., Azuma,T., and Nishioka,I. (1967). Tetrahydrocannabinolic acid, a genuine substance of tetrahydrocannabinol. Chem Pharm Bull (Tokyo) *15*, 1075-1076.

Yao, R. Biosynthesis of cuticular alkylresorcinols in selected grass species *Brachypodium distachyon* and *Secale cereale*. 2011. The University of British Columbia.

Ylihonko,K., Tuikkanen,J., Jussila,S., Cong,L., and Mantsala,P. (1996). A gene cluster involved in nogalamycin biosynthesis from *Streptomyces nogalater*: sequence analysis and complementation of early-block mutations in the anthracycline pathway. Mol Gen Genet *251*, 113-120.

Zdobnov,E. and Apweiler,R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847-848.

Zhan,J. (2009). Biosynthesis of bacterial aromatic polyketides. Curr Top Med Chem *9*, 1958-610.

Zhang,H., Matsuda,H., Yamashita,C., Nakamura,S., and Yoshikawa,M. (2009). Hydrangeic acid from the processed leaves of *Hydrangea macrophylla* var. thunbergii as a new type of anti-diabetic compound. Eur J Pharmacol *606*, 255-261.

Zhang,W. and Tang,Y. (2009). In vitro analysis of type II polyketide synthase. Methods Enzymol *459*, 367-393.

Zhang,W., Wilke,B.I., Zhan,J., Watanabe,K., Boddy,C.N., and Tang,Y. (2007). A new mechanism for benzopyrone formation in aromatic polyketide biosynthesis. J Am Chem Soc *129*, 9304-9305.

Zhou,H., Li,Y., and Tang,Y. (2010). Cyclization of aromatic polyketides from bacteria and fungi. Nat Prod Rep *27*, 839-868.

Zhu,B., Xiong,A.S., Peng,R.H., Xu,J., Zhou,J., Xu,J.T., Jin,X.F., Zhang,Y., Hou,X.L., and Yao,Q.H. (2008). Heat stress protection in Aspen sp1 transgenic *Arabidopsis thaliana*. BMB Rep *41*, 382-387.

Zuardi,A.W. (2006). History of cannabis as a medicine: a review. Rev Bras Psiquiatr *28*, 153-157.

APPENDIX

APPENDIX A

Determining the concentration of PDAL, HTAL, OA, and olivetol.

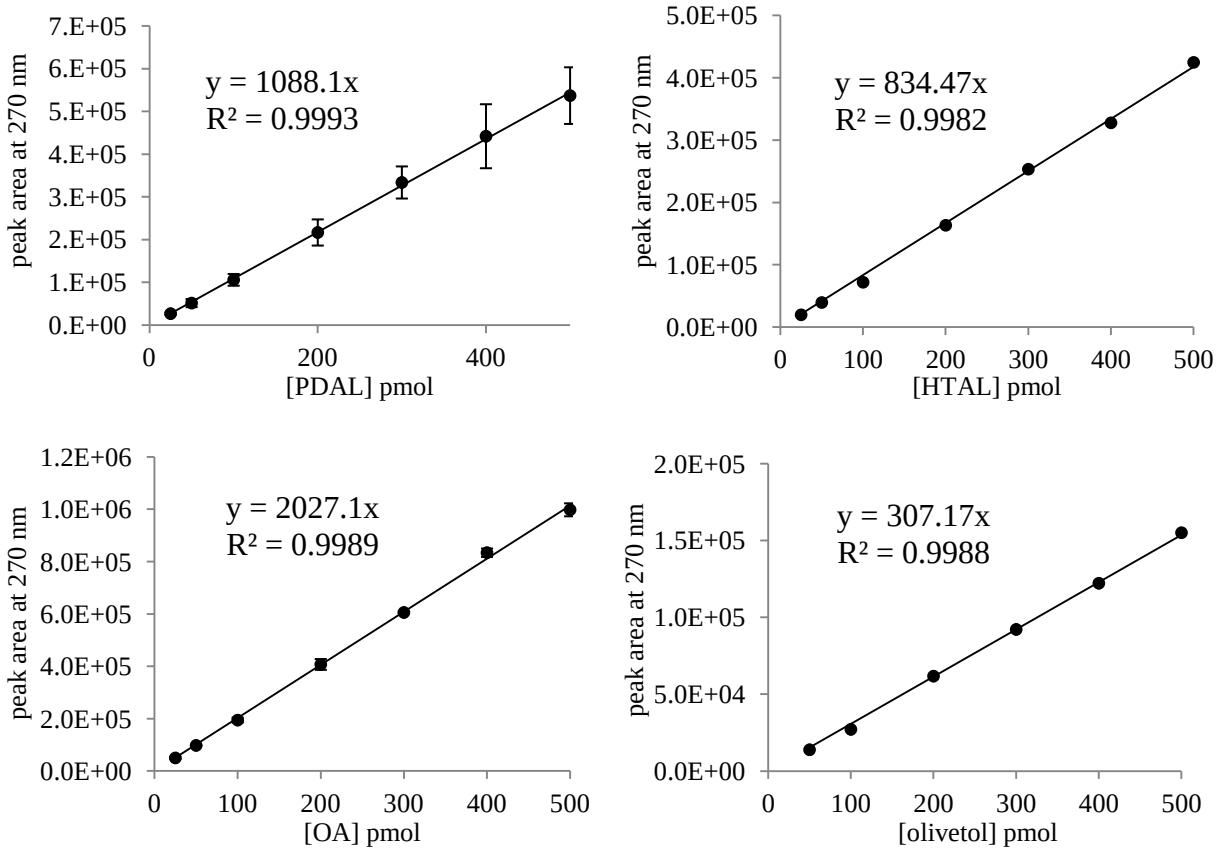


Figure A1 Standard curves used for product determination. Error bars are the range in duplicate samples. Equations are the linear regression lines with correlation coefficient expressed as R2.

APPENDIX B

Table B1 Injection and bleed times for the generation of antibodies targetting TKS

Description	Date	Day #
Pre-immune and 1° injection	1-Jun-2009	0
2° Injection	22-Jun-2009	21
1° bleed	2-Jul-2009	31
3° Injection	13-Jul-2009	43
2° bleed	23-Jul-2009	52
4° Injection	10-Aug-2009	70
3° bleed / Final 80 days	20-Aug-2009	80
5° injection (facultative)	7-Sep-2009	98
4° bleed	17-Sep-2009	108
6° injection (facultative)	5-Oct-2009	126
5° bleed	15-Oct-2009	136
7° injection (facultative)	2-Nov-2009	154
6° bleed / Final 164 day	12-Nov-2009	164
8° injection (facultative)	30-Nov-2009	182
7° bleed	10-Dec-2009	192
9° injection (facultative)	28-Dec-2009	210
8° bleed	7-Jan-2010	220
10° injection (facultative)	25-Jan-2010	238
9° bleed / Final 248 days	4-Feb-2010	248

APPENDIX C

Q-TOF MS/MS Analysis of an unknown protein.

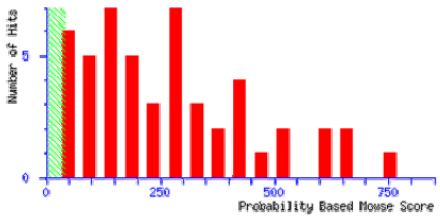
The identification of an unknown 60 kDa protein was also purified from both recombinant and trichome extracts, and sent to the PBI mass lab for Q-TOF MS/MS analysis. The unknown protein(s) was identified as an A. thaliana chaperonin homolog (annotated NP_175945) (Section 3.2.4.2).

Mascot Search Results

```
User            : NRC Mascot Daemon
Search title    : PLGS processed pkl files searched on MASCOT, Ottawa
MS data file    : \\DATA\MSPCDATA\MassSpec\Laboratory\Instruments\Q-ToF\PKL_files-2010\100204CS2_SG_03.pkl
Database        : NCBInr 04012010 (10274250 sequences; 3505793397 residues)
Taxonomy        : Viridiplantae (Green Plants) (729512 sequences)
Timestamp       : 10 Feb 2010 at 17:24:43 GMT
Protein hits    : gi|225435794 PREDICTED: hypothetical protein [Vitis vinifera]
                  gi|2506277   RecName: Full=RuBisCO large subunit-binding protein subunit beta, chloroplastic; AltName: Full=60 kDa cha
                  gi|15222729  CPN60B (CHAPERONIN 60 BETA); ATP binding / protein binding [Arabidopsis thaliana]
                  gi|147819511 hypothetical protein [Vitis vinifera]
                  gi|255564820 rubisco subunit binding-protein beta subunit, rubb, putative [Ricinus communis]
                  gi|3790441   chaperonin 60 alpha subunit [Canavalia lineata]
                  gi|2493650   RecName: Full=RuBisCO large subunit-binding protein subunit beta, chloroplastic; AltName: Full=60 kDa cha
                  gi|224104681 predicted protein [Populus trichocarpa]
                  gi|115443643 Os02g0102900 [Oryza sativa (japonica cultivar-group)]
                  gi|464727    RecName: Full=RuBisCO large subunit-binding protein subunit alpha, chloroplastic; AltName: Full=60 kDa ch
                  gi|223948025 unknown [Zea mays]
                  gi|242090109 hypothetical protein SORBIDRAFT_09g014430 [Sorghum bicolor]
                  gi|31193919  putative rubisco subunit binding-protein alpha subunit precursor (60 kDa chaperonin alpha subunit) [Oryza
                  gi|217074850 unknown [Medicago truncatula]
                  gi|242032147 hypothetical protein SORBIDRAFT_01g000380 [Sorghum bicolor]
                  gi|168056654 predicted protein [Physcomitrella patens subsp. patens]
                  gi|22250     mitochondrial chaperonin-60 [Zea mays]
                  gi|223635800 RecName: Full=(-)-limonene synthase, chloroplastic; AltName: Full=(-)-(4S)-limonene synthase; AltName: Fu
                  gi|2924773   mitochondrial chaperonin (HSP60) [Arabidopsis thaliana]
                  gi|255554262 chaperonin-60kD, ch60, putative [Ricinus communis]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 41 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



APPENDIX D

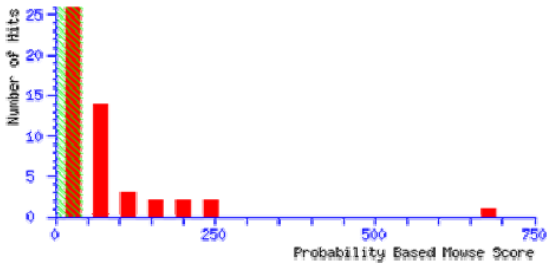
The identification of TKS was also confirmed by Q-TOF MS/MS analysis (Section 3.2.4.2).

Mascot Search Results

User : NRC Mascot Daemon
Search title : PLGS processed pkl files searched on MASCOT, Ottawa
MS data file : \\DATA\MSPCDATA\MassSpec\Laboratory\Instruments\Q-ToF\PKL_files-2010\100204CS2_SG_04.pk
Database : NCBInr 04012010 (10274250 sequences; 3505793397 residues)
Taxonomy : Viridiplantae (Green Plants) (729512 sequences)
Timestamp : 10 Feb 2010 at 17:25:23 GMT
Protein hits : gi|171363647 olivetol synthase [Cannabis sativa]
gi|2078350 transaldolase [Solanum tuberosum]
gi|225425278 PREDICTED: hypothetical protein isoform 2 [Vitis vinifera]
gi|29468084 aspartate aminotransferase [Oryza sativa]
gi|217795376 transaldolase [Dimocarpus longan]
gi|129916 RecName: Full=Phosphoglycerate kinase, cytosolic
gi|21536853 phosphoglycerate kinase, putative [Arabidopsis thaliana]
gi|217072408 unknown [Medicago truncatula]
gi|225464999 PREDICTED: hypothetical protein [Vitis vinifera]
gi|2605932 aspartate aminotransferase [Lotus corniculatus]
gi|147833248 hypothetical protein [Vitis vinifera]
gi|15224592 ASP1 (ASPARTATE AMINOTRANSFERASE 1); L-aspartate:2-oxoglutarate aminotrans
gi|120666 RecName: Full=Glyceraldehyde-3-phosphate dehydrogenase, cytosolic
gi|710596 mitochondrial aspartate aminotransferase [Glycine max]
gi|115445217 Os02g0236000 [Oryza sativa (japonica cultivar-group)]
gi|224109060 predicted protein [Populus trichocarpa]
gi|225448085 PREDICTED: hypothetical protein [Vitis vinifera]
gi|221255555 glyceraldehyde 3-phosphate dehydrogenase [Populus balsamifera]
gi|729503 RecName: Full=Naringenin,2-oxoglutarate 3-dioxygenase; AltName: Full=Flava
gi|226438612 chalcone synthase A [Rhabdothamnus solandri]

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 40 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



APPENDIX E

A total of 93 OAC-like sequences (including OAC) were aligned using ClustalW provided in Mega5 (Tamura et al., 2011) with parameters: Pairwise Alignment: Gap Opening Penalty 10; Gap Extension Penalty 0.1; Multiple Alignment: Gap Opening Penalty 10; Gap Extension Penalty 0.2; Protein Weight Matrix Gonnet; Residue-specific Penalties ON; Hydrophobic Penalties ON; Gap Separation Distance 4; End Gap Separation OFF; Use Negative Matrix OFF; Delay Divergent Cutoff (%) 30%.

Phylogenetic Relationship Analysis used the following parameters: Phylogeny Reconstruction; Statistical Method: Maximum Likelihood; Test of Phylogeny: Bootstrap method; Number of Bootstrap Replications: 1000; Substitution Model: Amino acid; Model/Method: Jones-Taylor-Thornton (JTT) + Gamma model; Rates among Sites: Uniform rates; Gaps/Missing Data Treatment: Complete deletion; ML Heuristic Method: Nearest-Neighbor-Interchange (NNI); Initial Tree for ML: Make initial tree automatically.

APPENDIX F

Sequences used during the phylogenetic analysis of OAC.

>A. coerulea 027_00390

MGDLKHMVLVKFKEDVVVEDILKGMEKLASEVDSIKSFEWGQDLESEEMLRQGFTHVFTMTFG
SSEDFAACQTHPSHVEFSTTFSAAIEKILLFHFPAVLVKAPAA

>A. coerulea 029_00097

MEEAKGVVKHVLMAKFKDEITPDRIDLLIKGYANLVPLIPPMKAFHWGKDVSIENLHQGFTHVF
ESTFESLEGVAEYISHPAHVEFATEFLTALDKVIVIDYKPTIANH

>A. lyrata 941765

MATSGFKHLVVVKFKEDAKVDEILEGLENLVSQIDTVKSFEWGEDKESHEMLRQGFTHAFSMTF
ENKDGYVAFTSHPLHVEFSAAFTAVIDKIVLLDFPVAAVKSSVVVTP

>A. thaliana AT3G17210

MAEAKGAVVKHVLLAKFKDGVSPETIEELIKGYANLVNLIEPMKAFHWGEDVSIENLHQGYTHI
FESTFETKEAVAEYIAHPAHVEFATIFLGSLDKALVIDYKPTSVSL

>A. thaliana AT5G22580

MATSGFKHLVVVKFKEDTKVDEILKGLENLVSQIDTVKSFEWGEDKESHDMLRQGFTHAFSMTF
ENKDGYVAFTSHPLHVEFSAAFTAVIDKIVLLDFPVAAVKSSVVATP

>A. thaliana TA34447_3702

GSKMEEAKGPVKHVLLASFKDGVSPEKIEELIKGYANLVNLIEPMKAFHWGKDVSIENLHQGYT
HIFESTFESKEAVAEYIAHPAHVEFATIFLGSLDKVLVIDYKPTSVSL

>B. distachyon 4g04380

MAGSGVVKHILLARFKDEVTPERLDQLIRGYAGLVAVVPSMKAFHWGTDVSIENMHQGFTHVF
ESTFESTEGVKEYIEHPNHVDFATEFLALAEKVIIIDYKPTAAT

>B. distachyon 4g42625

MGEFKHLCLVRFKEGVVVDNIIQELSKLATELDTVKYFGWGKDVLEQEALTQGFTHVFSMSFAT
AGDLAACMAHEKHAAFAATFMAALDKVVVMDFPFVTVKPAPEPEPAT

>B. rapa 021232

MEEAKGPVKHVLLAKFKDDVTPEKIDELIKGYANLVNLIEPMKAFHWGKDVSIENLHQGFTHIF
ESTFDSKEAVAEYVAHPIHVEFANMFLGSLDKVLVIDYKPTSV

>B. rapa Bra002422

MATSGFKHLVVVKFKEDAKVDEILKGLENLVSQIDSVKSFEWGEDNESHEMLRQGFTHAFSMTF
ENKDAYVSFTGHPLHVEFSAAFTAVIDKIVVMDFTVAAVKSPVVVAP

>B. rapa Bra020198

MAASEFKHLVVVKFKEDAKADEILKGSENLVSQIDSVKSFEWGEDKESHEMLRQGFTHAFSMTF
ENKDAYVSFTKHPLHVKFSAAFTAVIEKIVILDFTVAAVKSPVVVPP

>C. annuum TA6827_4072

NFASLINTATEQSRRRVGEMEGTKGGVVKHILLAKFKDEISPDKIDQLIKQYANLVNLIEPMKAS
HWGKDVSTENLQQGFTHVFESTFDSTEGVAEYVSHPVHVEFANLMLPQLEKVLVIDYKPEKVRP
LN

>C. clementina 025762m

MKAETKGRDMEEAKGVVKHVLLAKFKEGTAQDQIDQLIKDYANLVNLIEPMKSFQWGKDVSIE
NLHQGFTHIFESTFESTEGVAEYVAHPAHVEYANLFLANLEKVLVIDYKPTTVRV

>C. clementina 025899m

MGEGEEAAMGEFKHLVIVKFKEGVVVEDIVKGMKKLVSEIDAVKSFEWGQDVEGQEMLRQGF
THAFLMTFNKKEDYTTFASHPSHVEFSATFSAAIEKIVLLDFPTVLGKAPAA

>C. papaya 146.67

MEEAKGVVKHVLVAKFKEETLPAQIDELIKGYANLVNLIEPMKSFHWGKDVSIENLHQGFTHVF
ESAFESTEGIAEYIDHPAHVEFANLFLPVLEKFVVIDYKPTTVSF

>C. rubella 10014967m

MEEANKGPVKHILLAKFKDGVSPEKIEELIKGYANLVNLIEPMKAFHWGEDVSIENLHQGYTHIF
ESTFESKEAVAEYVAHPVHVEFATMFLGSLDKVLVIDYKPTSVSL

>C. rubella v10002301m

MATSGFKHLVVVKFKEDTKVDEILKGLENLVSQIDTVKSFEWGEDKESHEMLRQGFTHAFSMNF
ENKDGYVAFTSHPLHVEFSAAFTAVIDKIVVLDFPVAAVKSSVAV

>C. sativa OAC

MAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFE
SVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK

>C. sativa PK00183

MEKANKGVVKHLVIMKFKDGISDDQIEQMNKEYANLLNLVPSMKTLQLGKVVEMSPGNYKHG
NGGYTHIFESTFESMEGVAEYTFHPAHLHLGHLYSHTFDKVLVFDYIIPITTIS

>C. sativa PK17532

MGEANKGVVKHVFILKMKEGLSNDQIEQMNKDYANLVNLVPSMKALQWGKLEVNNKLGNGG
YTHIFESTFESMEGVAEYADHPAHLHLRNLYFHTLDKFLVFDYKPTIVLPNSSY

>C. sativa PK18164

MEEAKGVVKHVLLAKFKEGTSDDQIQQLIKGYANLLNLIPSMKSFHWGKDVSFENLHQGFTHIF
ESTFENTEGVAEYVAHPAHVEFANVFLSNLDKVVVFDYKPTTVLLP

>C. sativa PK28464

MEEAKGVVKHIVLLKFKDGITDDEIQQFINAYANIVNKIPVMKDFNWGKDVSCVNLHGGFTHIL
ASTFESIEDVAKYVHHPVHVEFGNLYHHNLEKFLIFDYKPTIFLP

>C. sativus 199990

MGEFKHLVVAKFKEGLNVDEIVAQVEKMVSDIDSVKSFEWGHDVEGQDMLTQGFTHVFSMTF
DDKEAITSFLTHPKHLEFCPTFSAAIDKIVVLDFPSILVKAAVPPPSPQPPPALEETPEAAVPPPSPQP
PPALEETPEAALPPVLTSTPLPTPATAAAATPASASA

>C. sativus 252360

MGDSKGVVKHILLAKFKNGISEEQIDQLIKGYANLVNLIEPMKAFQCGKDVSIENLHQGFTHVFE
STFESTQGIAEYISHPAHVEFANKFLSSLDKVIVIDYKPTLICN

>C. sativus 385330

MGEEFKHLVLVKFKEGAAVEEILKGMENLVSEVEAVKSFEWGQEMEGPEMLTQGFTHAFLMTF
ENKEAHTAFVGHPKHVEFSATFSAAIDKIVVLDFPAIRAKPPLPA

>C. sinensis 1g033664m

MGEGEEAAMGEFKHLVIVKFKEGVVVEDIVKGMKKLVSEIDAVKSFEWGQDVEGQEMLRQGF
THAFLMTFNKKEDYTTFASHPSHVEFSATFSAAIEKIVLLDFPTVLGKAPAA

>C. sinensis 1g033887m

MEEAKGVVKHVLLAKFKEGTAQDQIDQLIKDYANLVNLIEPMKSFQWGKNVSIENLHQGFTHIF
ESTFESTEGVAEYVAHPAHVEYANLFLANLEKVLVIDYKPTTVRV

>E. grandis B02216

MEEAKGLVRRVHLVKFKDGTPLARIEEILKNYANLVNLIEPLKSWHMGKNVSIIYLDQGFTDVFE
VTFESAEGLAEYADHPIHTEFSKQTRPYLDKVIAITYEPTV-CT

>E. grandis B02218

MEEAKGLAKRVLLVKFKDGTPPARLEEILKNYANLVNLVEPLKSWRMGKNVSFMNLDQGFTH
VFELTFESAKGLAEYADHPAHTEFSSQCRPYLDKFVVINYEPTVFCT

>E. grandis B03989

MEEASRPKEKTMEEAKGFVRRVLLFKFKAETAPDHIEQLIKDHSNLVNLVESLKSWHMGKDVS
VKYSNEGFTHVIELTFESTEGIAAYMGHPAHLEMHERIWPHLEKIVVIDYKPTLCN

>G. arboreum BM360640

ALREVRLMEEAKGVVKHILLAKFKDEIPPDKIEELIKGYANLVNLIHPMKAFHWGKDVSIENLHQ
GFTHVFESTFEST

>G. hirsutum TA22147_3635

LALREVRLMEEAKGVVKHILLAKFKDEIPPEKIEELIKGYANLVNLIHPMKAFHWGKDVSIENLH
QGFTHVFESTFESTEGIAEYIAHPAHVEFANLFLPSLDKVIVFDYKPTVVRC

>G. max 04g08730

MEEAKGVVKHVLLAKFKDDVTPERIEELIKDYANLVNLIPPMKSFHWGKDVSAENLHQGFTHVF
ESTFESPEGVAEYVAHPAHVEYANLFLSCLEKVIVIDYKPTVVKL

>G. max 06g08830

MEEAKGVVKHVFLAKFKDDVTPERIEELIKDYANLVNLIPPMKSFHWGKDVSAENMHQGFTHV
FESTFESTEGIAEYTAHPAHVEYATLFLSYLDKVIVIDYKPTIVKL

>G. max 09g23640

MGTFNHYVIVKLKDGVAVDELIQGLEKMVSGIDHVKSFEWGKDIESHDMLRQGFTHDFLMAFN
RKEEFNAFQTHLTHLEFTRVFSPAIEKIVVLDFPSNLVKAPA

>G. max 13g11380

MGTFNHYVIVKFNDGVAVDELIQGLEKMVSGIDHVKSFERGKDIESHDMLRQGFTHVFLMAFN
GKEEFNAFQTHVNDLEFTGLFSPAIEKIVVLDFPSNLMKAPA

>G. raimondii TA16529_29730

LALREVRLMEEAKGVVKHILLAKFKDEIPPEKIEELIKGYANLVNLIHPMKAFHWGKDVSIENLH
QGFTHVFESTFESTEGIAEYIAHPAHVEFANLFLPSLDKVIVFDYKPTVVRC

>G. soja TA2928_3848

EEEMEEAKGVVKHVLLAKFKDDVTPERIEELIKDYANLVNLIPPMKSFHWGKDVSAENLHQGFT
HVFESTFESPEGVAEYVAHPAHVEYANLFLSCLEKVIVIDYKPTVVKL

>H. lupus CL734Contig1

MEEAKGVVKHVLLAKFKEGIPDDQIEQLIKGYANLVNLIPPMKSFHWGKDVSFENLHQGFTHIFE
STFETTEGVAEYVAHPAHVEFANVFLSVLDKVVVFDYKPTTVLP

>L. usitatissimum 10037823

MEEGKGEVKHVLLAKFKPEISGEEIEKLIKDYANLVTLIEPMKAFHWGTDVSIENMHQGFTHIFV
STFESTEGIAEYVAHPVHVEFANLFLAALDKVLVFDYKPTSVK

>L. usitatissimum Lus10009407

MAAPEFKHLAIIKEGVVVDEMIKGMEKLVSEISLIKSLEWGQDLEGHEMLTQGFTHAFTLTFNSK
EDYIAYDNHPTHLEYAAAFSASIEKFVVLNYPTVLLSQT

>L. usitatissimum Lus10020555

MAAPEFKHLAIIKFKEGVVVDEMIKGMEKLVSQISLIKSLEWGQDLEGHEMLTQGFTHAFTLTFN
SKEDYIAYENHPTHLEYAAGFSASIEKFVVLNYPTVLLKQT

>M. acuminata 6T20510

MAVKHVLLAKFKEEVSPEAVDDLIKGYANLVSLITPMKAFHWGRDVSIENLHQGFTHVFESTFD
GVEGIAEYIAHPAHVEFANKFLPALEKVIVIDYAPTAVN

>M. acuminata 9T09110

MGSIRHLVLAKFKDEAAVEELLQGLQKLVSEIDVVKSFEWGEDVLKDERLGQGFTHAFLLTFGS
AEDLATYIKHPSHVAFGKAFRAAIDKILAIDFPVVTNEISA

>M. domestica MDP0000213440

MEEAKGVVKHVVLAKFKDGIAENKIEEIIKGFANLVNLIEPMKSFNWGKDVSSENLHQGFTHIFE
STFESTEGVAEYVAHPAHVDFANLFLSHLEKVIVIDYKPTAVSV

>M. domestica MDP0000528167

MEEAKGGELKNILMAKFKEGTSESQIEQLIESFANLVNLVEPMKSFHWGKELSIEKEEEGYIHVFE
TTFESVEGMAEYAVHPAHHDFAKLFLPNLEKMCSIGYRLVRRRDGTGQDGTGRDGTEQRFRVM
FGTSKMEQNGTGLND

>M. esculenta 019718m

MEEAKGVVKHVLLAKFKEGISSDEIEKLIKGYANLVNLIEPMKAFHWGTDVSIENLHQGFTHVFE
STFESTEGVAEYISHPAHVDFANLFLPALEKVVVIDYKPTTVRL

>M. esculenta 019824m

MGFEHLVIVRFKEGAVVADIMKGLEKLVSDVDLVKTFAWGEDLESPEMLTQGFTHAISMTFDK
KQDYTAFQTHPSHVEFSATFSAAIEKIVVLCFSSVQVKPAIA

>M. esculenta 029781m

MAEAKGVAKHIVLVKFKEEVSAEEIEQLIKGYANLVNLIPPMKSFRWGIDAGIENLTQGYTHAFE
STFETTQGIAEYVAHPAHIEFSNQLAPALHQVLTINYQSNTVHL

>M. esculenta 030990m

MGGYKHLVIFRFKAGAEVEEIMKGLEKLVSETDLVKSFEWGRDLDNPEILTQEFTHAISMTFDKK
QDYDAFQNLPSHIDFSLTFSTAVQKIVVLCFPSVQVKPAVA

>M. guttatus mgv1a016735m

MATEFKHLVLVKFKNDVVVGDVMKGLEKLVLEMDTVKSFVWGEDIESHEMLRQGFTHAFLMT
FGSKEDFTAFTSHPTHVDFSSTFSTAVDKAVLLDFPSVTVKPPPAAA

>M. guttatus mgv1a016769m

MEEGEVKHIVLAKFKESVSEEEIQESIKQYANLVNLVPSMKAFSWGKDLSKENMQQGFTHVFES
TFESTQGVAEYVSHPSHVEYATLLLSQLDNVIVVDYKPTPVRL

>M. guttatus mgv1a016804m

MEGEVKHILLAKFKQGITEDQIEEYIKQYANLVNLVPSMKSFSWGKDVSIENLNEGYTHIFESTFE
STQGIADYISHPDHVQYANTLLPQLEKVLVVDYKPTKVHL

>M. guttatus mgv1a024900m

GKHEVKHMLLAKLKQGISEDEIEGYIKQYANLVNLVPSMKAFAWGKDVSIENLNDGFTHVFEST
FASTQGIAEYLSDPDHVAFANTLLPQLDKVLVVDYMPTKVKL

>M. truncatula 019717m

MEEAKGVVKHIFLAKFKEEISSDQIEKLIKGYANLVNLIEPMKAFHWGTDVSNENLHQGFTHVFE
STFESTEGVAEYVSHPAHVEFANLFLAAAEKVIVIDYKPTTVRL

>M. truncatula 019718m

MEEAKGVVKHVLLAKFKEGISSDEIEKLIKGYANLVNLIEPMKAFHWGTDVSIENLHQGFTHVFE
STFESTEGVAEYISHPAHVDFANLFLPALEKVVVIDYKPTTVRL

>M. truncatula 019720m

MEEAKGVVKHIFLAKFKEEISSDQIEKLIKGYANLVNLIEPMKAFHWGTDVSNENLHQGFTHVFE
STFESTEGVAEYVSHPAHVEFANLFLAAAEKVIVIDYKPTTVRL

>M. truncatula 019721m

MEEAKGVVKHIFLAKFKEEISSDQIEKLIKGYANLVNLIEPMKAFHWGTDVSNENLHQGFTHVFE
STFESTEGVAEYVSHPAHVEFANLFLAAAEKVIVIDYKPTTVRL

>N. benthamiana CK296232

FGMEGGKVKHILLAKFKDGIPADQIDQLIKQYANLVNLIEPMKAFHWGENVSIENFHQGFTHVFE
STFDSTEGIAEYIDHPAHVEYANTLLPQLEKVLVIDYKPEKVGP

>N. benthamiana CK296233

FGMEGGKVKHILLAKFKDGIPADQIDQLIKQYANLVNLIEPMKAFHWGENVSIENFHQGFTHVFE
STFDSTEGIAEYIDHPAHVEYANTLLPQLEKVLVIDYKPEKVGP

>N. benthamiana EH366633

FGMEGGKVKHILLAKFKDGIPADQIDQLIKQYANLVNLIEPMKAFHWGENVSIQNFHQGFTHVF
ESTFDSTEGIAEYIDHPAHVEYANTLLPQLEKVLVIDYKPEKVGP

>N. tabacum EB446961

FGMDAAKGGGVVKHILLAEFKDGIPADQIDQLIKQYANLVNLIEPMKAFHWGENVSIENFHQGF
THVFESTFDSTEGIAEYIDHPTHVEYANILLPQLEKVLVIDYKPEKVGP

>O. sativa Os01g33160

MAAETPAAGRSGVLKHIVLARFKEEVTPERLDHLIRGFGGLVNLVPSMKAFNWGTDVSIENMHH
GFTHVFECTFESTEGVKEYIEHPAHLEFAKEILLAMEKTLIIDYMPTAVNNS

>O. sativa Os11g05290

MGEVKHLCLVKFKEEVLVDDILQGMTKLVSEMDMVKSFEWGKDVLNQEMLTQGFTHVFSLTF
ASSEDLTTYMSHERHQEFAGTFMAAIDKVVVVDFPVVASKPAPPAAPAAAA

>P. patens Pp1s189_106V6

MGEVKHLCLVKFKEGVVVEDVLKGMTDLVAGMDMVKSFEWGQDVLNQEMLTQGFTHVFSLT
FAFADDLATYMGHDRHAAFAATFMAALDKVVVIDFPVVVAKPPPPA

>P. patens Pp1s85_83V6

MSGVVKHILLLKFKDSISPDEQTALIQKYAALPNSISAMKGFEWGTDISVENMHQGFTHAFISTFD
SPQGRDEYLVHPVHDSFAKELLGAVDNALVFDFNPTVALKANS

>P. persica ppa013300m

MGSAVDLGNNNMGEFKHLVIVKFKEDVVVDDILKGLENLVAEIDAVKSFEWGQDLESQELLRQ
GFTHVFLMTFDKKDDYAVFQSHPKHQEFSAIFSTVIEKLVLLDFPPTLVKTPPKAAPPEAPPEAPP
A

>P. persica ppa013682m

MEEAKGVVKHVLLAKFKEGISENQIEELTKGYAKLVDLIEPMKSFHWGKDVSIENLHQGFTHIFE
STFESMEGVTEYVAHPAHVDFANLFLSHVEKVIVFDYKPTMVRV

>P. trichocarpa 0010s04670

MPRCSMRLVQWKKNVKKASSLVFKKPDSLSVGRNLEETKTESKRERERERDTVLMEEAK-G-
VVKHVLLAKFKEGIPSDEIEKLIKGYANLVNLIEPMKAFEWGTDVSIENMHQGFTHIFESTCESKE
GIAEYIGHPAHVDFGGLFLPALEKVIVFDYKPTVARF

>P. trichocarpa 0010s04700

MSKRRPAWFSSRNLEETKTESKRERDTVVMEEAKGVVKHVLLAKFKEGIPSDEIEKLIKGYANL
VNLIEPMKAFEWGTDVSIENMHQGFTHIFESTFESKEGIAEYIGHPAHVDFGGLFLPALEKVIVFD
YKPTVARF

>P. trichocarpa 0010s16070

MEQVKGVVKHIVLAKFKEGVTEEEIEKHIKDYANLLNHIEHMKSFHWGTDVSIENLHQGFTHIFE
ITFETLDGRSAYVAHPAHVDFGTAFLTILEKIVVVDFVPTLVKL

>P. trichocarpa 0010s16080

MEGGKGVVKHIVLAKFKEGTTEEEIEKHIKNYADLLNHIEHMKSFEWGTDVSIENLHQGFTHIFE
ATFETLEGRSAYVAHPAHVKFGTALFPTLEKVIVFDYVPKA

>P. vulgaris 091002521m

MEEAKGVVKHIVLAKFKDDITAEKIEELIKGYANLVNLVPPMKSFHWGTDVSAENLHQGFTHVF
ESTFESTEGVAEYVAHPAHVEFANLFISSTEKVIVIDYKPTI

>P. vulgaris 091014673m

MRACCIGLCIYTWLLGTTFIHKSCVVAMGDFNHYVVVKFKDGVAVEELIQGLEKMVSGIHEVKS
FEWGKDIESHDMLRQGFTHVFLMTFNGKEEFNTFRTHPNHIEFSGVFSPSIEKIVVLDFPSKLVKA
PA

>P. vulgaris 091024159m

MLAMGEFKHFVIVKFKEGVAVDELTKGMEKLVSEIDAVKAFEWGRDIESLDVLRQGFTHAFLM
TFNKKEDFAAFQSHPNHVEFSTTFSAAIENIVLLDFPSTLVKAPA

>R. communis 29889.m003356

MAEEVKEVHHLVMAKFKEDITPEKIEELIEGYANLVNLVPPMKAFHWGTDVSIENLNEGFTHIFH
ATIESREGLAEYVNHPAHIEFGNLMTPAMEKVIVVDFSPSLVQLQRE

>R. communis 29988.t000006

MAGFKHLVIVKFKEDAAVEEIVKGMEKLVSESDLVKSFAWGQDSEGPEMLTQGFTHAFSMTFD
KKEDYTAFQTHPNHVEYSATFSAAIEKIVVLFFPYVQVKATASA

>R. communis 30170.m013937

MEEAKGVVKHVLLAKFKDEIPTHEIDKLIKGYANLVNLIPPMKAFHWGSDVSIENLHQGFTHVFE
SSFESTEDIAEYISHPAHVEFANLFLRAVDKVIVIDYKPTIVHL

>S. bicolor 03g027060.1

MASGGVVKHILLARFKEDVTQERLDELIRGYAALVAAVPSMKAFHWGTDVSIENMHQGFTHVF
ESTFESTEGIKEYIEHPKHVEFANVFLPVLDKVIVIDYKPTSVN

>S. italic 003420m

MAAGGGSGVVKHIVLARFKEEVTPERLDQLIQGYAALVDAVPSMKSFHWGTNVSIENTHEGFT
HVFESTFESTEGIKEYVKHPAHVEYANEFLSVSEKILIIDYKPTSVN

>S. italica Si011410m

MAEVKHLCLVKFKEGVVVEDILKGMADLAAQMDMVKSFEWGKDVLNQEMLTQGFTHVFSLT
FASADDLTAYMAHEKHAAFAATFMAALEKVVVIDFPVVIAKPPPQA

>S. italica Si027006m

MAPRSIALALRSSLQSHSECGMGEVKHLCLVKFKEGVVVEDVLKGMADLAAQMDMVKSFEWG
QDVLNQEMLTQGFTHVFSLTFASADDLTAYVSHEKHAAFAATFMAALEKVVVIDFPVAIAKPPP
QA

>S. lycopersicum 05g053310

MEGGKGGVVKHILLAKFKDGIPPEQIDQLIKQYANLVNLVEPMKAFQWGKDVSIENLHQGFTHV
FESTFDSLEGVAEYIAHPVHVEYANTLLPQLEKFLIVDYKPQ

>S. lycopersicum 11g066950

MANEFKHLVLVKFKEDVVVEDILKELEKLVQEMDIVKSFVWGKDVESHEMLRQGFTHAIIMTF
NSKEDYQTFANHPNHVGFSATFATVIDKAVLLDFTAISGKAT

>S. moellendorffii 98687

MAQENPKSGLKHVLLARFKPEISEEEVASLIAGYEALPQSIEAMKGFEWGTDVSVENLHQDFTH
VFTSTFESPEGRDAYLVHPAHTAYANKLLPALDKVIVLDFHPKVVCHRS

>S. tuberosum TA38774_4113

SRKKSEKIEMEGGKGGVVKHILLAKFKDGIPPEQIDQLIKQYANLVNLVEPMKAFHWGKDVSIE
NLHQGFTHVFESTFDSLEGVAEYVAHPVHVEFANTMLPQLEKVLIIDYKPQ

>T. cacao Tc05_g028060

MVLIKRKHVGPSQNQTQIKPSKFKDETTAEKIEELIKGYANLVNLIEPMKAFQWGKDVSIENLHQ
GFTHVFESTFESTEGIAEYVAHPVHVQFANLFLGHLEKVLVIDYKPTIARC

>T. halophilav10015069m

MATSGFKHLVVVKFKEDAKIDEILKGLENLVSQIDSVKSFEWGEDKESHEMLRQGFTHAFSMTF
ENKDAYVSFTSHPLHVEFSTAFTAAIDKIVLLDFTVAAVKSPVEASP

>V. vinifera GSVIVP00011995001

MEEAKGLVKHVLLAKFKDSTPPDQIEELIKGYANLVNLVPPMKAFHWGKDVSIENMHQGFTHV
FESTFESVEGIAEYVSHPAHVEFANLFLPHLEKVIVFDYKPITVHL

>V. vinifera GSVIVP00019670001

MCGQLKRVEEGKGVVKHVLLAKFKDSTPPDQIEELIKSYANLVSLIPPMKAFHWGKDVSIENMH
QGFTHVFESTFESVEGMAEYVSHPAHVEAANRFLPHLEKVIVLDYKPTAVHL

>V. vinifera GSVIVP00033194001

MIFYKEIASFFYIYNREEGKGVVKHVLLAKFKDSTPPDQIEELIKSYTNLVSLIPPMKAFHWGKDV
SIENMHQGFTHVFESTFESVEGMAEYVSHPAHVEAANRFLPHLEKVIVLDYKPTAVHL

>Z. mays GRMZM2G050730

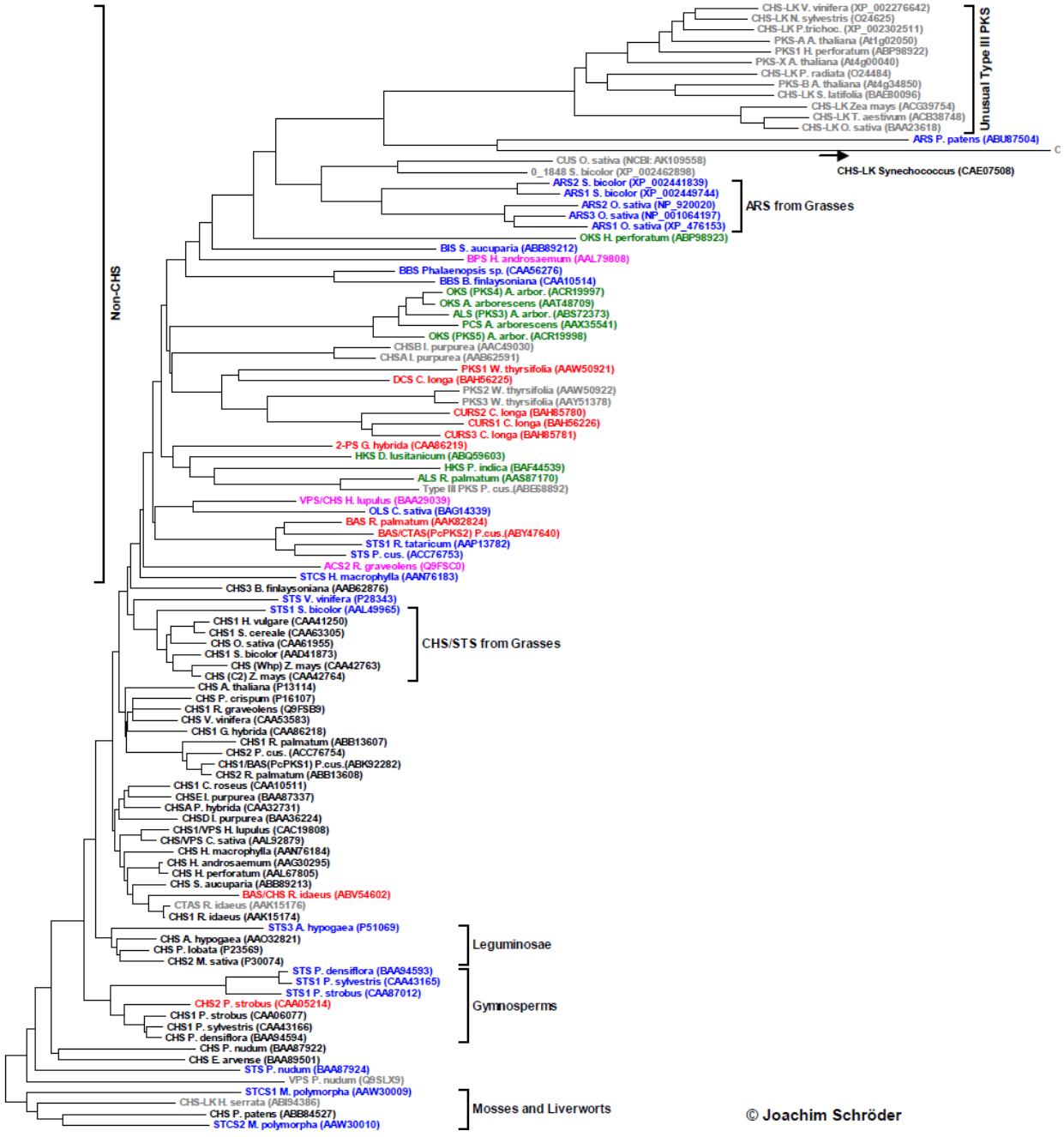
MGEVKHLCLAKFKDGVVVDDVLKGMTDLAAGMDTVKSFEWGQDVLNQEMLTQGFTHVFSLT
FASAADLAAYMAHDSHTAFAATFMAAIDKVLVVDFPVVVAKPPPPA

>Z. mays GRMZM2G351990

MPTASSRRTFSEAVSTDDEEAARGGKVGRPSPPRCSSRRGTDVSIENMHQGFTHVFESTFESTEGI
KEYIEHPAHVEFASVFLPVLEKVLIIDYKPTSAN

APPENDIX G

The Relationship of Type III PKS by J. Schröder (2010). Available at http://www.biologie.uni-freiburg.de/data/bio2/schroeder/Plant_TypeIII_Tree.pdf



Some notes

- **Color scheme**
  Red: one or two condensations
  Blue: three condensations, aldol condensation type cyclization to end product (e.g. STS)
  Black: **CHS**, three condensations, Claisen condensation type cyclization to end product
  Violet: other substrates as CHS, but three condensations, Claisen condensation type cyclization to end product
  Green: more than three condensations
  Grey: several definitions:
  - *in vitro* function unknown or not found so far
  - *in vitro* function known, but not physiological role (e.g. products unknown from that plant)

- **Note**
  The page '**Links to Sequences of the Relationship Tree**' contains
  - Explanations to names and sequences
  - Links to the protein sequences
  - Links to the pages discussing the enzymes

- **The tree does not contain all type III PKS from plants.**
  This applies most importantly to CHSs:
  - Chalcone synthase (CHS) is the prototype for these proteins for at least two reasons:
    a) it was the first member identified (*Petroselinum crispum*),
    b) it appears to be ubiquitous in plants,
  - Nevertheless, CHSs are in most cases only included if they are of special interest, e.g.
    a) for historical reasons,
    b) the first 3D-structure,
    c) if the plant also contains non-CHS type III enzymes,
  - CHSs are in many plants encoded in gene families, for closely related proteins with CHS activities (if the members were tested for function at all). Usually in such cases the tree contains only one representative member, to avoid cluttering with closely related proteins

- **The tree was developed with the program TREECON for Windows** (Van de Peer and De Wachter, 1994), using the inbuilt matrix for amino acid sequences, and the neighbour-joining method for distance calculations (Saitou and Nei, 1987).

References

- Saitou, N., Nei, M., 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. Molecular and Biological Evolution 4, 406-425.
- Van de Peer, Y., De Wachter, R., 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Computational Applications in Biosciences 10, 569-570.